UNIVERSIDAD AUTÓNOMA DE MADRID

Programa de Doctorado en Biociencias Moleculares

Fidelity of Human Immunodeficiency Viruses Type 1 and Type 2 Reverse Transcriptases in DNA Synthesis Reactions using DNA and RNA Templates

Alba Sebastián Martín

Madrid, noviembre, 2018

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

Programa de doctorado en Biociencias Moleculares

Fidelity of Human Immunodeficiency Viruses Type 1 and Type 2 Reverse Transcriptases in DNA Synthesis Reactions using DNA and RNA Templates

Memoria presentada por Alba Sebastián Martín, graduada en Biología, para optar al título de doctora en Biociencias Moleculares por la Universidad Autónoma de Madrid

Director de la Tesis: Dr. Luis Menéndez Arias

Este trabajo ha sido realizado en el Centro de Biología Molecular 'Severo Ochoa' (UAM-CSIC), con el apoyo de una beca de Formación de Profesorado Universitario, financiada por el Ministerio de Educación, Cultura y Deporte (FPU13/00693).

Abbreviations

3TC	2’ 3’-dideoxy-3’-thiacytidine
AIDS	Acquired immunodeficiency syndrome
AMV	Avian myeloblastosis virus
APOBEC	Apolipoprotein B mRNA editing enzyme
ATP	Adenosine 5’ triphosphate
AZT	3’-azido-2’, 3’-dideoxythymidine (zidovudine)
AZT-MP	3´-azido-2´, 3´-dideoxythymidine monophosphate
AZTppppA	3´azido-3´-deoxythymidine-(5´)-tetraphospho-(5´)-adenosine
bp	Base pair
BSA	Bovine serum albumin
CA	Capsid protein
cDNA	Complementary DNA
Cir-Seq	Circular sequencing
CypA	Cyclophilin A
dATP	2’-deoxyadenoside 5’-triphosphate
dCTP	2’-deoxycytidine 5’-triphosphate
ddC	2’, 3’-dideoxycytidine
ddI	2’, 3’-dideoxyinosine
dGTP	2’-deoxyguanoside 5’-triphosphate
DNA	Deoxyribonucleic acid
dNTP	2’-deoxynucleoside 5’-triphosphate
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
dTTP	2’-deoxythymidine 5’-triphosphate
EDTA	Ethylenediaminetetraacetic acid
Env	Envelope proteins
ESCRT	Endosomal sorting complex required for transport
FeLV	Feline leukemia virus
f_{ext}	Mispair extension ratio
f_{ins}	Misinsertion efficiency constant
FIV	Feline immunodeficiency virus
FPCM	Fundación Parque Científico de Madrid
gRNA	Genomic RNA
HiRes-Seq	High resolution sequencing
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2

HTLV-III	Human T-lymphotropic virus type III
ICTV	International Committee on Taxonomy of Viruses
IN	Integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k_{cat}	Steady-state turnover number
K_d	Dissociation equilibrium constant
K_m	Michaelis constant
k_{obs}	Apparent kinetic constant
k_{off}	Dissociation rate constant
k_{pol}	Nucleotide incorporation rate
k_{pol}/K_d	Catalytic efficiency
lacZ	Gene encoding the β-galactosidase enzyme
LAV	Lymphoadenopathy-associated virus
LEDGF	Lens epithelium-derived growth factor
LTR	Long-terminal repeat
MA	Matrix protein
MCS	Multicloning site
MLV	Murine leukemia virus
mRNA	Messenger RNA
NC	Nucleocapsid protein
Nef	Negative regulating factor
NGS	Next-generation sequencing
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFV	Prototype primate foamy virus
PIC	Pre-integration complex
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
PPT	Polypurine tract
PR	Protease
RBS	Ribosome binding site
Rep-Seq	Replicated sequencing
RF	Replicative form
RNA	Ribonucleic acid
RNase H	Ribonuclease H
RNAP	RNA polymerase
RT	Reverse transcriptase
SAMHD1	Sterile α motif domain and histidine-aspartic domain-containing protein 1

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIV	Simian immunodeficiency virus
SSC	Saline-sodium citrate
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SU	Surface subunit
TAM	Thymidine analogue mutation
TAM1	Thymidine analogue mutation pathway 1
TAM2	Thymidine analogue mutation pathway 2
TM	Transmembrane subunit
TRIM5α	Tripartite motif-containing protein 5
T/P	Template/primer
U3	Unique 3'-region
U5	Unique 5'-region
Vif	Viral infectivity factor
Vpr	Viral protein R
Vpu	Viral protein unique
Vpx	Viral protein X
WT	Wild-type
(w/v)	(weight/volume)
X-Gal	5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside
XMRV	Xenotropic murine leukemia virus-related virus

Amino acids, one and three letters codes

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

INDEX

Summary

Human immunodeficiency viruses type 1 and 2 (HIV-1 and HIV-2, respectively) are the causative agent of acquired immunodeficiency syndrome (AIDS). Most of the available clinically approved antiretroviral drugs directed against HIV inhibit the viral reverse transcriptase (RT), a multifunctional enzyme with RNA- and DNA-dependent DNA polymerase activity, as well as ribonuclease H, strand-transfer, and strand displacement activities. Nucleoside RT inhibitors (NRTIs) constitute the backbone of current therapies against both types of HIV, although the acquisition of NRTI resistance is mediated by the development of different mutational pathways. In HIV-2, resistance to all approved nucleoside analogues is conferred by the combination of RT substitutions K65R, Q151M and M184V. Interestingly, in highly divergent HIV-1 RTs, K65R alone confers >8-fold increased accuracy of DNA synthesis. In this Thesis, we have determined the intrinsic fidelity of DNA synthesis of wild-type (WT) HIV-2 group A (strain ROD) RT and mutants K65R and K65R/Q151M/M184V. Our results show that in HIV-2_{ROD} RT those changes have a relatively small impact on nucleotide selectivity. Furthermore, we found that there were less than 2-fold differences in error rates obtained with forward mutation assays using mutant and WT HIV-2_{ROD} RTs. A different conformation of the β3-β4 hairpin loop in HIV-1 and HIV-2 RTs could probably explain the different effects of K65R.

In addition to its importance in natural infection, RTs are also widely used in biotechnology for their capacity to synthesize complementary DNA using RNA as templates. Intrinsic fidelity of RT's DNA-dependent DNA synthesis has been extensively studied. In M13mp2 *lacZα* forward mutation assays, WT HIV-1 RTs of group M/subtype B showed >10-fold higher error rates than murine leukemia virus (MLV) and avian myeloblastosis virus (AMV) RTs. This Thesis aimed to determine error rates for RNA-dependent DNA synthesis catalyzed by several RTs, including WT HIV-1_{BH10}, HIV-1_{ESP49}, AMV and MLV RTs, and the high-fidelity mutants of HIV-1_{ESP49} RT K65R and K65R/V75I. Our results show that these retroviral RTs have less than 2-fold differences in fidelity, with error rates in the range of 2.5×10^{-5} and 3.5×10^{-5}. These results were consistent with the existence of a transcriptional inaccuracy threshold, generated by the RNA polymerase while synthesizing the RNA template used in the assay. A modest but consistent reduction of the inaccuracy threshold was achieved by lowering the pH and Mg^{2+} concentration of the transcription reaction, needed to synthesize the RNA template. Despite assay limitations, we concluded that HIV-1_{BH10} and HIV-1_{ESP49} RTs are less accurate when copying DNA templates than RNA templates. Analysis of the RNA-dependent mutational spectra revealed a higher tendency to introduce large deletions at the initiation of reverse transcription by all HIV-1 RTs except the double-mutant K65R/V75I. Data provided by deep sequencing experiments also showed similar differences in accuracy between HIV-1_{BH10} and K65R/V75I$_{ESP49}$ RTs.

Resumen

El síndrome de la inmunodeficiencia adquirida (SIDA) está causado por el virus de la inmunodeficiencia humana de tipo 1 y 2 (VIH-1 y VIH-2, respectivamente). La mayoría de los fármacos antirretrovirales aprobados clínicamente para su uso contra la infección por el VIH actúan sobre la retrotranscriptasa (RT) viral. La RT es una enzima multifuncional con actividad ADN polimerasa dependiente de moldes ARN y ADN, así como con actividad ribonucleasa H. Los inhibidores de la RT análogos a nucleósido constituyen el eje central de las terapias actuales contra ambos tipos de VIH. En la RT del VIH-2, la combinación de cambios de aminoácido K65R, Q151M y M184V confiere resistencia a todos los análogos a nucleósido aprobados. Cabe destacar que en RTs de VIH-1 muy divergentes, la mutación K65R confiere por sí misma >8 veces mayor fidelidad en la síntesis de ADN. En esta Tesis, hemos determinado la fidelidad intrínseca para la síntesis de ADN dependiente de moldes de ADN de la RT de tipo salvaje (WT) del VIH-2, grupo A, cepa ROD y de las RTs mutantes K65R y K65R/Q151M/M184V. Nuestros resultados mostraron que estos cambios tienen un efecto relativamente pequeño (menos de 2 veces) en la selectividad de nucleótido de la RT del VIH-2_{ROD}. Una conformación diferente de la horquilla β3-β4 en las RTs del VIH-1 y VIH-2 probablemente podría explicar los diferentes efectos de K65R.

Además de su importancia en la infección natural, las RTs también se utilizan ampliamente en biotecnología por su capacidad de sintetizar ADN complementario utilizando ARN como molde. Se ha estudiado ampliamente la fidelidad intrínseca de la síntesis de ADN dependiente de ADN. En ensayos genéticos basados en la expresión del gen *lac*Zα del fago M13mp2, las RTs WT del VIH-1, grupo M, subtipo B mostraron tasas de error >10 veces mayores que el virus de la leucemia de ratón (MLV) y el virus de la mieloblastosis de aves (AMV). En esta Tesis se han determinado las tasas de error de síntesis de ADN dependiente de ARN para varias RTs, incluyendo las de los virus WT del VIH-1_{BH10}, VIH-1_{ESP49}, AMV y MLV, y las RTs de alta fidelidad del VIH-1_{ESP49} con los cambios K65R y K65R/V75I. Nuestros resultados muestran diferencias de menos de 2 veces en la fidelidad de estas RTs, con valores de tasas de error entre 2.5×10^{-5} y 3.5×10^{-5}. Estos resultados fueron consistentes con la existencia de un umbral de error de la transcripción, generado por la ARN polimerasa al sintetizar el ARN utilizado como molde en el ensayo. Se logró una reducción modesta pero consistente del umbral de error al disminuir el pH y la concentración de Mg^{2+} de la reacción de transcripción. A pesar de las limitaciones del ensayo, concluimos que las RTs del VIH-1_{BH10} y VIH-1_{ESP49} son menos precisas cuando se copian moldes de ADN que de ARN. El análisis de los espectros de mutación dependientes de ARN reveló una mayor tendencia a introducir deleciones largas al inicio de la retrotranscripción en todas las RTs del VIH-1, excepto la del doble mutante K65R/V75I. Los datos proporcionados por secuenciación masiva muestran también diferencias similares en la fidelidad entre las RTs del VIH-1_{BH10} y del doble mutante K65R/V75I$_{ESP49}$.

1. INTRODUCTION

1.1. Human immunodeficiency virus

Reverse-transcribing viruses are characterized by the presence of a monophyletic hallmark enzyme named reverse transcriptase (RT). In 2017, the International Committee on Taxonomy of Viruses (ICTV) regrouped most of these viruses into the order *Ortervirales* (*Orter* is an inversion of "retro," which derived from reverse transcription). Orterviruses share several features, such as (i) a homologous aspartate protease domain in their polymerase polyprotein, (ii) the use of host tRNA molecules as primers for genome replication, and (iii) the two principal protein components of the virions (capsid and nucleocapsid) (reviewed in Krupovic *et al*., 2018). Within the *Ortervirales* order, human immunodeficiency viruses (HIV) belong to the *Retroviridae* family, which naturally infects a variety of animals including non-human primates. Taxonomically, HIV is classified in the *Orthoretrovirinae* subfamily and the *Lentivirus* genus, in which is considered the type species (https://talk.ictvonline.org/taxonomy/).

HIV is classified into two different types, HIV-1 and HIV-2. HIV-1 was firstly described in 1983 in the Luc Montagnier's laboratory of the Institute Pasteur, and was initially named lymphadenopathy-associated virus (LAV) (Barré-Sinoussi *et al*., 1983). HIV-1 was also referred to as human T-lymphotropic virus III (HTLV-III) (Gallo *et al*., 1984; Levy *et al*., 1984; Popovic *et al*., 1984). HIV-2 was later described in 1986 at the Institute Pasteur (Clavel *et al*., 1986). Both viruses are considered the etiological agents of the acquired immunodeficiency syndrome (AIDS), but they differ in their virological and clinical manifestations.

Like HIV-1, HIV-2 is transmitted perinatally, through several sexual practices, by contact with infected blood products and with contaminated syringes during injectable drug use. However, HIV-2 is generally less pathogenic and infectious, showing lower rates of sexual and vertical transmission, longer asymptomatic phases (ten years or more) and a slower progression to AIDS in comparison with HIV-1. In addition, long-term non-progressors are more frequent in people living with HIV-2 than in those infected with HIV-1 (86-95% *versus* 5-15%, respectively). These patients maintain high CD4+ T cell counts over many years in the absence of antiretroviral therapy, despite a possible low but detectable viremia. Although HIV-2 infection courses with less morbidity and mortality, it has fewer and less effective treatment options than HIV-1 infections, yielding a much poorer immunological recovery after starting treatment (reviewed in Menéndez-Arias and Álvarez, 2014; Saleh *et al*., 2017).

HIV originated from distinct zoonotic transmissions of simian immunodeficiency viruses (SIV) from primate to human. HIV-1 came from at least four independent events that leaded to the emergence of groups M (main), O (outlier), N (non-M, non-O) and P (putative) (**Figure 1**). Viruses belonging to groups M and N originated from SIV strains found in chimpanzees (SIVcpz, *Pan*

troglodytes troglodytes), while those SIV strains originating groups O and P were found in gorillas (SIVgor, *Gorilla gorilla gorilla*), in central Africa (D'arc *et al*., 2015). Likewise, HIV-1 group M can be subdivided into 11 subtypes from A to K. On the other hand, HIV-2 has independently crossed into humans at least nine times (groups A to I) from SIV strains of sooty mangabey monkeys (SIVsmm, *Cercocebus atys atys*) in West Africa. No subtypes have been formally described but some preliminary data suggest that HIV-2 group A may be divided in two different subtypes with distinct geographical origins (reviewed in Ndung'u and Weiss, 2012; Visseaux *et al*., 2016). HIV-1 presents multiple circulating recombinant forms, while HIV-2 only has two recombinant forms (one circulating referred to as CRF01_AB, and one unique) (Visseaux *et al*., 2016).

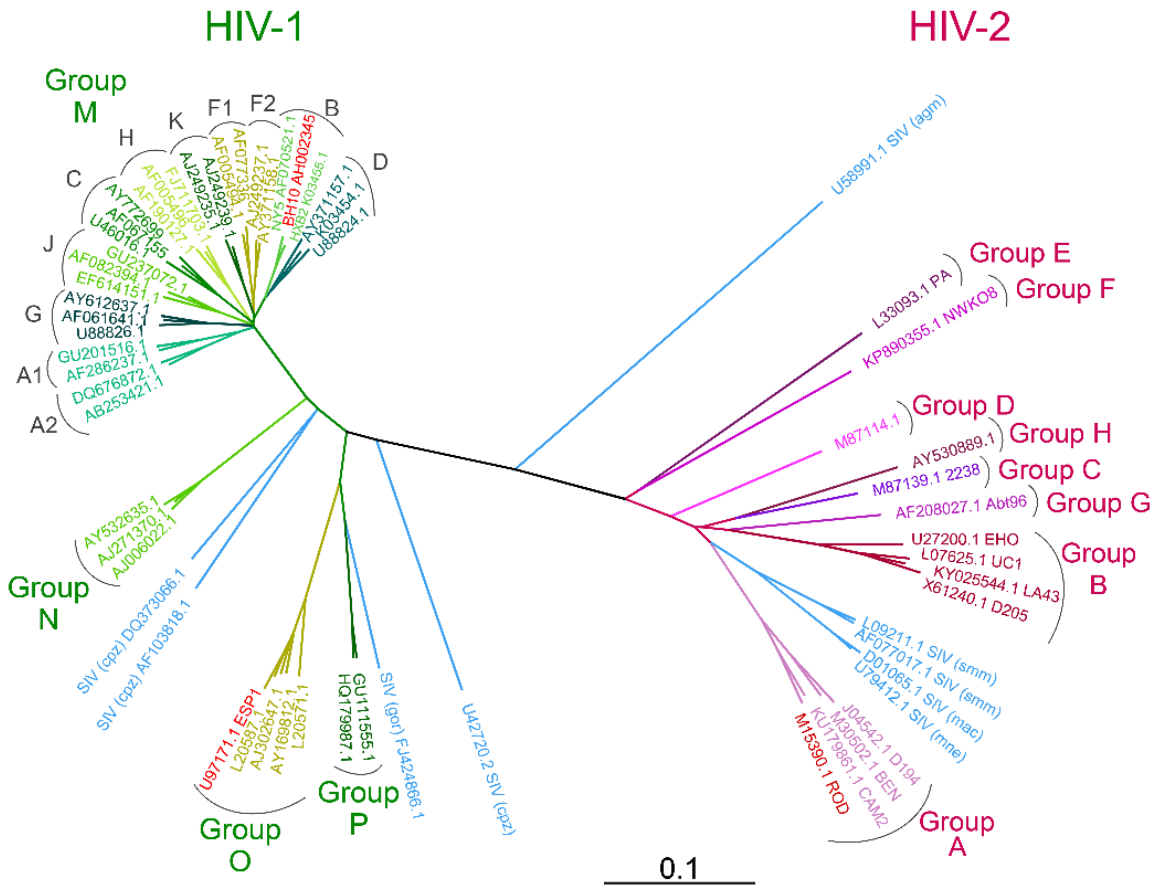


Figure 1. Phylogenetic tree analysis of the *pol* gene of HIV reference strains. Sequences are identified by their GenBank accession codes, from where they were retrieved. Phylogenetic analyses were carried out by using the Geneious 6.1.8. software (https://www.geneious.com), and assuming the sequence alignment provided by the ClustalW algorithm (Thompson *et al*., 1994). Evolutionary relationships were inferred using the neighbor-joining method, constructing the tree to scale (represented in the line below), with branch lengths in the same units as those of the evolutionary distances. The Tamura-Nei method (Tamura and Nei, 1993) was used to compute evolutionary distances, whose units indicate the number of base substitutions per site. Strains of HIV classified in diverse subtypes and groups are indicated for HIV-1 (green) and HIV-2 (fuchsia). Strains of SIV (simian immunodeficiency virus) that infect chimpanzees (cpz), gorillas (gor), sooty mangabeys (smm), African green monkeys (agm), macaques (mac) and pigtail macaques (mne) are represented in blue. Well-characterized reference strains are colored in red.

Despite the high diversity, pandemic HIV infection is mainly caused by HIV-1 group M viruses and to a lesser extent by HIV-1 group O, together with HIV-2 groups A and B viruses. The original outbreak of HIV-1 occurred in Kinshasa around 1920 (Faria *et al*., 2014). Since the onset of the AIDS epidemic, HIV has infected around 76.6 million individuals and has claimed the lives of 35.2 million people (http://www.unaids.org/en). Nowadays, around 36.7 million people are living with HIV, of which between one and two million are HIV-2 infections.

1.2. HIV virion

The cell-free form of the virus is surrounded by a lipid bilayer derived from the plasma membrane of the infected cell, acquired during virion budding (**Figure 2A**). The viral envelope contains trimeric glycoproteins (Env) composed of transmembrane subunits (TM, gp41), non-covalently linked to surface subunits (SU, gp120). Associated with the inner leaflet of the envelope through lipid contacts, the matrix protein (MA, p17) organizes in a matrix layer. The core of the virion is made of capsid proteins (CA, p24) forming a conical shell and its content. The capsid cone is composed of roughly two hundred and fifty CA hexamers, and twelve CA pentamers asymmetrically distributed (seven at the wide end and five at the narrow end) (Ganser-Pornillos *et al*., 2012). This capsid contains the viral genome that is composed of two molecules of positive-sense single-stranded unspliced RNA ((+)ssRNA). The genomic RNA (gRNA) is stabilized by interactions with the nucleocapsid protein (NC, p7). Inside the virion, there are also three essential enzymes for HIV replication: (i) the reverse transcriptase (RT, p66/p51), responsible for synthesizing a double-stranded DNA (dsDNA) from the viral (+)ssRNA; (ii) the integrase (IN, p32) that catalyzes the insertion of the viral dsDNA into the host cell genome; and (iii) the protease (PR, p11), which carries out the proteolytic processing of precursor polyproteins, allowing the maturation of the virion (for reviews see Adamson and Freed, 2007; Ganser-Pornillos *et al*., 2012; Freed, 2015).

The gRNA contains a 5' cap and a 3'-polyadenylated tail, resembling a messenger RNA (mRNA). Elements structurally important for HIV replication are located within the 5' termini. After reverse transcription, proviral DNA is flanked in both extremes by long terminal repeats (LTRs) (**Figure 2B**), including regions U3 (unique 3')-R (repeat)-U5 (unique 5') (Menéndez-Arias *et al*., 2017). The HIV genome (around 9.7 kb) contains nine different open reading frames. Three major genes (*gag*, *pol* and *env*) code for the structural and enzymatic proteins described above. Two genes code for the regulatory proteins Tat (transactivator protein) and Rev (RNA splicing-regulator), essential for virus replication (Romani and Engelbrecht, 2009); whereas four additional genes code for accessory proteins known as Vif (viral infectivity factor), Nef (negative regulating factor), Vpr (viral protein R); and Vpu/Vpx (viral protein unique/X). Vpu is found exclusively in HIV-1 and Vpx

is present in HIV-2, although they are not functional homologues (Faust *et al*., 2017). Accessory proteins were dispensable in studies *in vitro*, although they are considered increasingly important for efficient infection *in vivo* (Seelamgari *et al*., 2004; Anderson and Hope, 2004; Richter *et al*., 2009). Vif, Vpr and Vpx are packaged into the virion (Yu *et al*., 1988; Cohen *et al*., 1990; Müller *et al*., 2000; Kao *et al*., 2003; González *et al*., 2017). Several cellular proteins (e.g., actin, cyclophilin A and multiple RNA-binding proteins) are also incorporated into the nascent virion, but their precise functions (if any) are unclear in most cases (reviewed in Briggs and Kräusslich, 2011; Telesnitsky and Wolin, 2016).

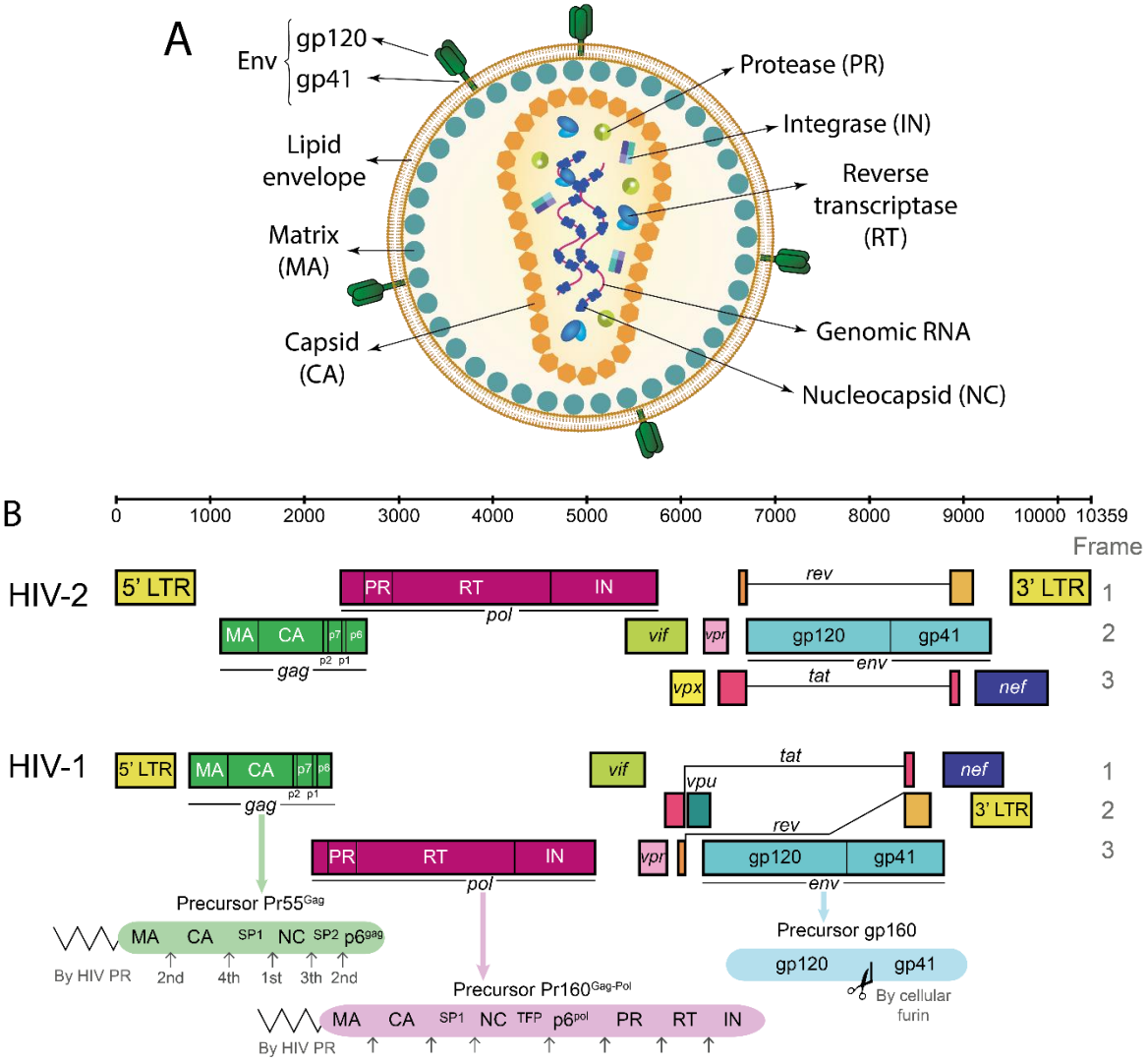


Figure 2. Structure and genome organization of HIV. (A) Schematic representation of the viral particle. **(B)** HIV-1 and HIV-2 genome organization, adapted from the HIV sequence compendium 2018 (Foley *et al*., 2018). Long-terminal repeats (LTRs) flanking the HIV genome, and open reading frames for major genes (*gag*, *pol*, *env*), regulatory genes (*tat*, *rev*), and accessory genes (*nef*, *vif*, *vpu*, *vpr/vpx*) are shown as rectangles, distributed in three lines following the corresponding frame (indicated at right). Numbering nucleotide positions are indicated above. Below, polyprotein products obtained from the three major genes. Arrows show the site of cleavage mediated by HIV PR (in the case of Gag and Gag-Pol precursors) or by the cellular furin (in the case of precursor gp160). The order of proteolytic processing is indicated for Gag. A myristic acid molecule is attached to the N-termini of Gag and Gag-Pol polyproteins.

1.3. Cycle of HIV infection

HIV-1 is adapted to replicate in human cells, in particular CD4+ T cells such as helper T-lymphocytes, macrophages and microglial cells. The replication cycle can be divided into early and late phases. The early phase encompasses the events that occur from virus binding to the host cell surface until the integration of the HIV-1 genome into that of the host cell. Viral infection initiates with the high-affinity binding of gp120 to the cellular surface receptor CD4 (**Figure 3**). This interaction produces a conformational change that allows the virus to interact with a co-receptor, such as CCR5 or CXCR4 (reviewed in Fassati, 2012). The initial attachment of the virus triggers rearrangements of gp41 protein, leading to the fusion of the viral envelope with the cell membrane, and delivering the viral core into the cytoplasm. The released viral core is transported towards the nucleus by cellular motor proteins, likely using microtubule and actin network for trafficking (reviewed in Suzuki and Craigie, 2007).

The dissociation of CA from the conical shell is referred to as uncoating, which may happen near the plasma membrane or close to the nuclear pore. Controversy relates to the high percentage of nonfunctional particles that hampers the correlation between phenotype and progression of the infection (reviewed in Ambrose and Aiken, 2014). Recent ground-breaking studies suggested that long-lived/gradually uncoating cores are more likely to enter the nucleus and establish infection than the cores undergoing early/abrupt uncoating in the cytoplasm (Mamede *et al*., 2017; Francis and Melikyan, 2018a; reviewed in Francis and Melikyan, 2018b). Thus, viral RNA would be reverse transcribed within the protected environment of the capsid, preventing the trigger of the innate immune sensing, while recruiting nucleotides for reverse transcription through a strongly electropositive pore found at the center of CA hexamers (Jacques *et al*., 2016). Indeed, reverse transcription itself could partly trigger the uncoating, destabilizing the core due to the pressure exerted by the size and rigidity of the nascent RNA/DNA hybrid or dsDNA genome (Hulme *et al*., 2011; Rankovic *et al*., 2017). After the stripping of most (if not all) CA monomers from the core, the viral dsDNA remains associated with proteins, forming the pre-integration complex (PIC).

HIV-1 PIC may contain Vpr, MA, IN, RT, NC and triple-stranded cDNA flap (generated after reverse transcription), as well as cellular cofactors such as the lens epithelium-derived growth factor (LEDGF/p75) or importin α, among others. Several cellular proteins regulate the uncoating (e.g., CypA and CPSF6) and help the PIC to break into the nucleus (e.g., Tnpo3, Nup358, and Nup153) (reviewed in Fassati, 2012; Suzuki and Craigie, 2007; Zhan *et al*., 2010). Once in the nucleus, HIV-1 PIC is directed towards chromatin via the interaction of the viral IN with LEDGF/p75 that is a host transcription factor, thereby stimulating targeted integration into the host genome (Cherepanov *et al*., 2003). After integration, the HIV-1 provirus is packaged into nucleosomes. A multifactorial context determines whether HIV transcription will be silenced (leading to latency) or become active

(resulting in a productive infection). This context refers to the chromatin landscape, nucleosome positioning and post-transcriptional modifications (reviewed in Cary *et al*., 2016; Ne *et al*., 2018).

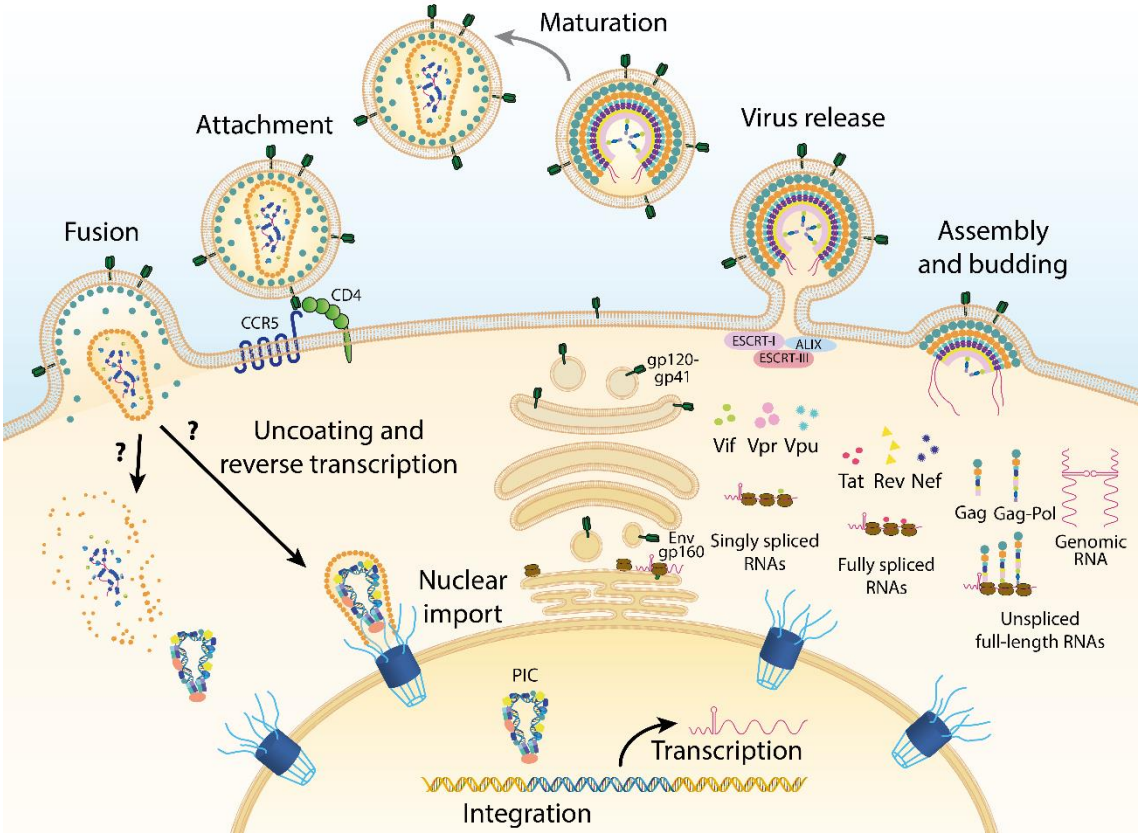


Figure 3. Schematic representation of the HIV infection cycle. Early phase starts with the attachment of the virus to the plasma membrane of the host cell, followed by fusion and entry. If uncoating occurs abruptly or gradually is not known. After reverse transcription, the PIC enters the nucleus by the nuclear pores, and the proviral DNA integrates into the host genome mediated by the viral IN. Transcription of the proviral DNA marks the beginning of the late phase, which includes expression of viral genes, assembly of viral particles, and their release and maturation into infectious viruses.

The late phase comprises events from gene expression to the release and maturation of new virions. HIV-1 transcription is controlled by a promoter located at the 5' LTR, and it initiates at one of the three consecutive guanosine residues found in the junction U3-R, by recruiting critical cellular factors (reviewed in Ne *et al*., 2018). However, at this stage, predominantly short incomplete viral transcripts are produced by human RNA polymerase (RNAP) II. Expression of viral Tat overcomes this block by recruiting a super elongation complex that phosphorylates the RNAP II on its carboxyl terminal, improving its processivity. This activation yields fully spliced RNAs that result in translation of only three viral proteins: Tat, Rev and Nef. Rev is then reimported through its nuclear localization signal, where it directs the export of intron-containing RNAs. Singly spliced messengers are translated into the remaining viral accessory proteins and the Env polyprotein; while unspliced RNAs provide Gag and Gag-Pol polyproteins, as well as gRNA molecules (**Figure 3**) (reviewed in Faust *et al*., 2017; Hidalgo and Swanson, 2017).

After translation in the cytosol, Gag seems to specifically bind a gRNA dimer, multimerize and reach the plasma membrane by a still-undefined pathway (reviewed in Freed, 2015; Dubois *et al*., 2018). Assembly disposes radially with the sequential Gag (or Gag-Pol) domains arranged with the N-terminus on the plasma membrane and the C-terminus pointing toward the center of the virion. This association is facilitated by the myristic acid moiety of MA (**Figure 2B**), and likely prompts the formation of lipid microdomains or rafts, where retrovirus assembly appears to be initiated. MA also seems to be responsible for the incorporation of mature Env glycoproteins, previously processed by the cellular protease furin. The principal sites of Gag-Gag interaction are found in the N- and C-termini of CA, as well as in the SP1 linker. Lastly, NC recruits the gRNA dimer, whereas the p6 domain adapts the host endosomal sorting complex required for transport (ESCRT) machinery, which catalyzes the membrane fission step to complete the release of new virions from the host cell surface (**Figure 3**) (reviewed in Ganser-Pornillos *et al*., 2012).

Virus particles bud in a noninfectious or immature form. Gag-Pol polyprotein, which arises from a ribosomal frameshift occurring with a frequency of around 5% during Gag translation, contains the viral PR. After dimerization, functional PR is liberated by autoproteolysis during or immediately after budding. Processing is likely initiated by the PR at the rim of the incompletely closed Gag shell and then proceeds as a "wave" into the interior of the lattice. Proteolysis proceeds at a specific rate and order (**Figure 2B**), resulting in the formation of structural and functional proteins: MA, that remains attached to the envelope; CA, that rearranges to form the conical shell; NC, which possesses nucleic acid chaperone activity and condenses with the RNA into a compact ribonucleoprotein, with a more stable RNA; p6, PR, RT and IN. These structural rearrangements lead to the formation of mature infectious virions (reviewed in Ganser-Pornillos *et al*., 2012).

Accessory proteins do not have enzymatic activity but interact with cellular restriction factors, facilitating viral replication (for a review see Faust *et al*., 2017). Vif counteracts the antiviral effects of host APOBEC3G by inducing its ubiquitinitation and subsequent degradation in cellular proteasomes. In the absence of Vif, APOBEC3G is packaged into virions and induces lethal mutagenesis via hypermutation of the viral genome by deaminating cytidine residues (reviewed in Okada and Iwatani, 2016). Vpu is unique to HIV-1 and acts on BST-2/Tetherin, which blocks virus release by anchoring virions to the plasma membrane. Vpu can inhibit the recycling of internalized BST-2 to the plasma membrane, or induce its ubiquitination. In HIV-2, the function of Vpu is exerted by the Env glycoprotein (reviewed in Strebel, 2013). In contrast, Vpx is found only in HIV-2 and targets SAMHD1 for proteasomal degradation, avoiding the depletion of cellular deoxynucleoside triphosphate (dNTP) pools to a level similar to non-dividing cells and resting T cells (Lim *et al*., 2012; Laguette *et al*., 2012). HIV-1 does not present a similar anti-SAMHD1 activity, but replication is not avoided due to a higher binding affinity of HIV-1 RT to dNTPs relative to other lentiviruses. Vpr has many reported functions, including LTR transactivation, nuclear import of the PIC, cellular

apoptosis, cell cycle arrest, and activation of the DNA damage response. Finally, Nef promotes HIV-1 infectivity, likely mediated by the downregulation of the integral membrane protein SERINC5.

1.4. HIV RT

1.4.1. Structure

HIV-1 RT is an asymmetric heterodimer composed of two related subunits of different lengths, p66 and p51 (di Marzo Veronese *et al*., 1986), which share a common amino terminus. Many crystal structures of the enzyme have been obtained under different conditions (e.g., apoenzyme alone, complexed with a variety of nucleic acids, with or without antiretroviral drugs) (for a review, see Menéndez-Arias *et al*., 2017). Analysis of X-ray crystallographic data revealed important structural features in the HIV-1 RT. The large subunit p66 (560 amino acids) is composed of two spatially distinct domains, DNA polymerase and ribonuclease H (RNase H). The DNA polymerase domain includes three subdomains called fingers (residues 1–85 and 118–155), palm (86–117 and 156–236), and thumb (237–318) (**Figure 4A**) (Kohlstaedt *et al*., 1992; Jacobo-Molina *et al*., 1993). A region known as connection (residues 319–426) links the DNA polymerase domain with the RNase H domain, which extends between residues 427 and 560 (Jacobo-Molina *et al*., 1993).

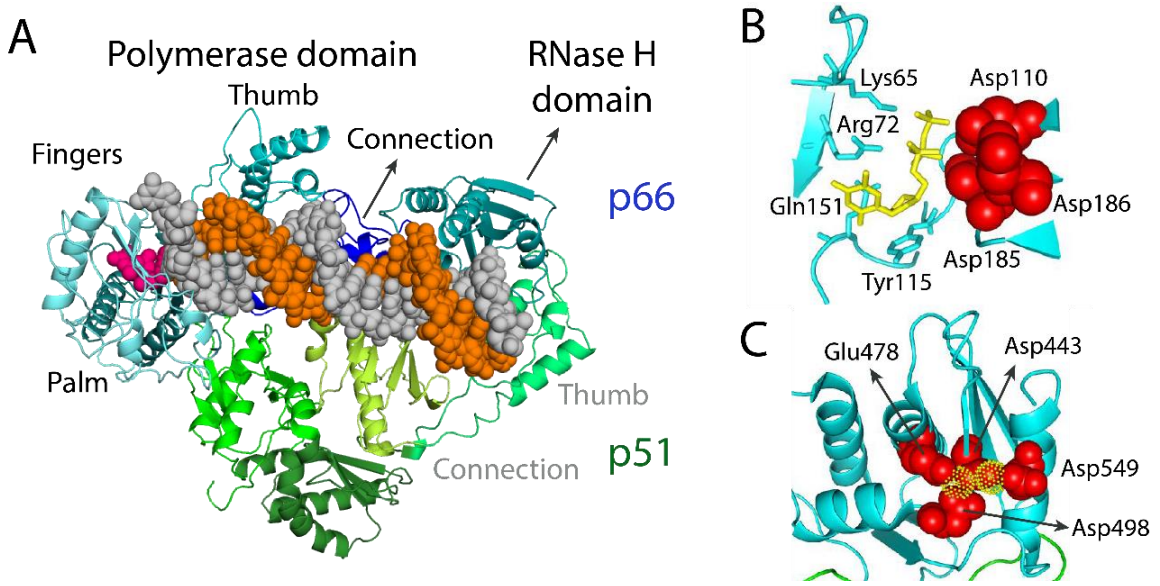


Figure 4. HIV-1 RT structure and active sites. (A) Crystallographic structure of HIV-1 RT bound to dsDNA and an incoming dNTP (PDB file 1RTD). Subunits p66 (blue) and p51 (green) are represented with ribbons, using different shades for each subdomain. Connection and thumb subdomains of p51 are inverted relative to those of p66. Template (grey), primer (orange) and the incoming dNTP (fuchsia) are represented with spheres. **(B)** Nucleotide binding site in the DNA polymerase domain showing the location of catalytic residues Asp110, Asp185 and Asp186 (red spheres), and the side-chains of Lys65, Arg72, Tyr115, and Gln151 (blue sticks). The incoming nucleotide is represented with yellow sticks. **(C)** RNase H active site showing the location of Asp443, Glu478, Asp498, and Asp549 (red spheres), and the coordinating metal ions (dot yellow surfaces). Figures 4B and 4C were adapted from Menéndez-Arias *et al*. (2017).

Meanwhile, the smaller subunit p51 (440 amino acids) is identical to the N-terminal portion of p66, lacks the RNase H domain, and folds into the same p66 polymerase subdomains. However, the spatial organization of the subdomains relative to each other is different in p66 and p51. Thus, p51 is more tightly packed, with the connection rotated to cover the palm and displacing the thumb (**Figure 4A**). As a result, p51 is catalytically inactive, while p66 contains the active sites for the enzymatic activities of RT (LeGrice *et al*., 1991; Hostomsky *et al*., 1992). For the most part, p51 has a structural role anchoring the proper folding of p66, and playing a pivotal role in stabilizing the heterodimer. This subunit also participates in binding the tRNA primer (Zakharova *et al*, 1995; Dufour *et al*., 1998) and has an impact on the RT's RNase H activity (Chung *et al*., 2013).

The nucleic acid binding cleft is formed by the fingers, palm and thumb subdomains of p66, which provide lateral and apical interactions with the nucleic acid substrate. In addition, the 'floor' of the cleft is shaped by the connection subdomains of both subunits and the thumb subdomain of p51 (Jacobo-Molina *et al*., 1993). The RNase H domain plays only a minor role in the template-primer binding. HIV-1 RT binds the nucleic acid substrate in a conformation that places the substrate in direct contact simultaneously with both the DNA polymerase and RNase H active sites. The distance between the two active sites is about 60 Å, equivalent to 17 nucleotides of a DNA/DNA complex or 18 nucleotides of a RNA/DNA hybrid (reviewed in Esposito *et al*., 2012).

The active site of the polymerase domain presents three highly conserved aspartate residues at positions 110, 185 and 186, located within the palm subdomain of p66 (**Figure 4B**). Asp185 and Asp186, together with Tyr183 and Met184, are part of the YXDD motif, highly conserved in retroviral RTs. The carboxylate moieties of the conserved aspartates bind two divalent ion cofactors (likely Mg^{2+}) that are essential for catalysis (Larder *et al*., 1987; Lowe *et al*., 1991). The catalytic metals in coordination with both the 3'-OH of the primer terminus, and the non-bridging oxygen of the α-phosphate of the incoming dNTP appear to be critical to maintain the appropriate distance for the nucleophilic attack (Huang *et al*., 1998; Mendieta *et al*., 2008).

The incoming dNTP is tightly coordinated by the side-chains of Arg72 and Lys65 in the fingers subdomain of p66, which bind the β- and γ-phosphates, respectively (**Figure 4B**). Other important residues that delineate the dNTP binding pocket are Ala113, Ala114, Tyr115 and Phe116. Residue Tyr115 contributes to the binding of the deoxyribose ring of the incoming dNTP and discriminates between deoxy and ribonucleoside triphosphates due to the presence of the 2'-OH in their ribose ring (Cases-González *et al*., 2000). Another remarkable interaction in the dNTP binding site is that of Gln151 with the 3'-OH of the incoming dNTP (reviewed in Sarafianos *et al*., 2009).

On the other hand, the RNase H domain folds into five β-stranded sheets and four α helices. The active site is surrounded by a cluster of four conserved acidic residues: Asp443, Glu478, Asp498 and Asp549 (motif DDE), which bind two divalent Mg^{2+} cations (**Figure 4C**). The two metal ions

are asymmetrically coordinated in the active site and have distinct roles: ion A activates the nucleophilic water molecule, and ion B (possibly in conjunction with metal ion A) stabilizes the transition state intermediate. Mutations in any of the Asp443, Glu478, and Asp498 residues abolish the enzyme activity (Mizrahi *et al*., 1990, 1994).

Unlike in the case of HIV-1 RT, there is only one crystal structure available for HIV-2 RT (PDB 1MU2) (Ren *et al*., 2002), corresponding with the unliganded form of the enzyme. Both polymerases share around 60% identity and show a similar organization into domain/subdomains, although HIV-2 RT heterodimers are more stable than those of HIV-1 (Divita *et al*., 1995). The large subunit of HIV-2 RT (p68) is 559 residues long, whereas the size of the smaller subunit p54 is uncertain. Biochemical evidence suggests that cleavage leading to the formation of p54 occurs after Met484, and not at the equivalent site of HIV-1 RT (i.e., between Phe440 and Tyr441) (Fan *et al*., 1995). This difference is likely due to the distinct specificities of the two HIV proteases. In the crystal structure, the smaller subunit contains only 427 residues, attributed to degradation of p68 subunit by bacterial proteases (Bird *et al*., 2003).

The most significant structural differences between HIV-2 and HIV-1 RTs are located in the non-nucleoside RT inhibitors (NNRTI)-binding pocket, near the DNA polymerase active site. The presence of Ile181 and Leu188 in HIV-2 RT, instead of Tyr as in HIV-1 RT, together with the lack of the aromatic side chains at these positions, abolish the effects of ring stacking interactions between the RT and many NNRTIs. These differences contribute to HIV-2 RT resistance to many NNRTIs (reviewed in Herschhorn and Hizi, 2010). In the absence of crystals of ternary complexes including HIV-2 RT, a template-primer and the incoming dNTP, any functional or mechanistic interpretation of the role of different residues relies on information available for HIV-1 RT.

1.4.2. RT activities

1.4.2.1. DNA polymerization

The RT is able to synthesize DNA using both RNA and DNA molecules, as templates and primers. In general, RTs may be able to incorporate a few hundred nucleotides in one round of DNA synthesis. Processivity and fidelity, among other features of the RT's DNA polymerization activity, have been widely optimized and exploited for their use in biotechnology, as discussed below. Despite the lack of proofreading 3'-5' exonuclease activity, HIV RTs are able to carry out the reversal of dNTP incorporation (i.e., pyrophosphorolysis), using pyrophosphate (PPi) or an NTP as the acceptor substrate (reviewed in Esposito *et al*., 2012). This capacity is particularly important for resistance mechanisms involving the removal of azidothymidine (AZT) and other nucleoside analogues (Arion *et al*., 1998; Meyer *et al*., 1999; reviewed in Menéndez-Arias, 2008).

Additionally, RTs can perform strand DNA displacement coupled to DNA polymerization. This property has been attributed to specific residues of the fingers subdomain (e.g., Phe61 in HIV-1 RT) (Fisher *et al*., 2003). Moreover, RTs can add non-templated nucleotides at the 3'-end of nascent DNA strands in reactions carried out *in vitro* (Patel and Preston, 1994; Golinelli and Hughes, 2002).

1.4.2.2. RNase H

The RNase H activity present in RTs hydrolyzes phosphodiester bonds in RNA strands when being part of RNA/DNA heteroduplexes, rendering RNAs with free 3'-OH and 5' phosphate ends. The RNase H activity is mediated by a nucleophilic attack of a water molecule on the scissile phosphate, breaking the bond between two adjacent ribonucleosides in the RNA strand. The RNase H specificity depends on the particular RNA minor groove width and its interaction with the "primer grip" (residues 227−235 of p66), a region that contacts the template-primer. The RNA/DNA hybrid has a minor groove width of ~9-10 Å, while non-hydrolysable polypurine tracks (PTTs) show a width of 7 Å probably due to the presence of A-tracts (reviewed in Schultz and Champoux, 2008; Esposito *et al*., 2012; Menéndez-Arias *et al*., 2017).

Three distinct types of cleavages have been reported based on the position of DNA polymerase and RNase H domains of the RT in relation to the RNA/DNA complexes (**Figure 5**).

- DNA 3' end-directed cleavage: this hydrolysis is polymerase-dependent and occurs when the 3' end of the growing DNA strand locates at the polymerase active site, annealed to a longer RNA. The catalysis results in the cleavage of RNA 15–20 nucleotides away from the DNA end.
- RNA 5' end-directed cleavage: the polymerase domain binds internally to a DNA strand annealed with the 5' end of a recessed RNA, cleaving it at 13−19 bases away. The 5' end is not recognized at a nick, but a gap of two or more bases upstream is sufficient for such recognition.
- Internal cleavage: in this type of cleavage, the nucleic acid strand-ends do not influence RT positioning. The hydrolysis occurs internally on a RNA/DNA hybrid.

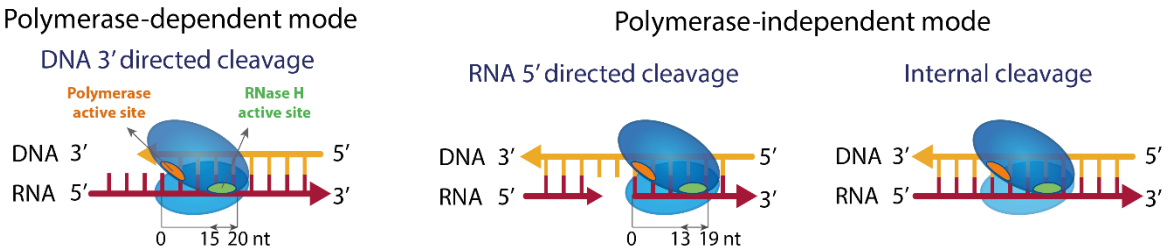


Figure 5. HIV RT binding modes to RNA/DNA hybrids and effects on RNase H cleavage. Polymerase-dependent or -independent binding produce different cleavage patterns, which result from different orientations of the RT relative to the template-primer. Adapted from Menéndez-Arias *et al*. (2017).

Nucleotide sequence in the vicinity of the cleavage site is a main determinant of RNase H specificity (Schultz *et al*., 2009; Kielpinski *et al*., 2017). The RNase H activity of HIV-2 RT is about ten times lower than that of HIV-1 RT, due to the presence of Gln294 in p54, instead of proline as occurs in the HIV-1 RT p51 subunit (Sevilya *et al*., 2001, 2003; Herzig and Hizi, 2015). RNase H activity plays an important role in strand transfer events by generating the single-stranded DNA (ssDNA) region that anneals to the acceptor molecule (Luczkowiak *et al*., 2018). This capacity of HIV RT to switch templates is based on sequence complementarity at the 3'-end of both donor and acceptor strands, and is facilitated by the presence of specific sequences in the viral genome (e.g., repeated sequences, and pausing sites). Strand transfer is assisted by the chaperone activity of the viral NC protein that enhances minus-strand transfer (Darlix *et al*., 2011). Reverse transcription involves at least two strand transfer events, although additional ones were suggested to occur at broken as well as intact templates (reviewed in Herschhorn and Hizi, 2010).

1.4.3. Fidelity of RTs

Genetic variation remains one of the major obstacles to the eradication of HIV, helping the virus to escape host immune responses, and to develop resistance to antiretroviral drugs (Megens and Laethem, 2013). Diversity emerges from high rates of replication, mutation and recombination, among others (Smyth *et al*., 2012). Mutations may be introduced into the viral genome by different enzymes throughout the replicative cycle: (i) RT during RNA- and DNA-dependent DNA synthesis, (ii) replicative host DNA polymerases (e.g., α, δ, γ, ε) involved in proviral DNA amplification, and (iii) the cellular transcriptional machinery (RNAP II) (reviewed in Menéndez-Arias, 2013a). The effect of host cell DNA polymerases on the viral mutation rate is expected to be minimal due to their high fidelity (estimated in 10^{-9} to 10^{-12} mutations/base pair/cycle) (reviewed in Svarovskaia *et al*., 2003). However, the contribution of RNA-dependent relative to DNA-dependent DNA synthesis by RTs is unknown, as well as that of human RNAP II during the synthesis of viral RNAs.

In addition to their role in viral infection, RTs play a fundamental role in supporting the biotechnological advances introduced in the late 1990s, such as cDNA microarrays and next-generation sequencing (NGS) technologies, due to their ability to synthesize DNA using RNA templates. For these purposes, wild-type (WT) and engineered RT variants of avian myeloblastosis virus (AMV), murine leukemia virus (MLV), HIV-1, and *Geobacillus stearothermophilus* group II introns have been developed into more efficient tools to study gene expression by increasing their catalytic efficiency, processivity, thermostability or fidelity of DNA synthesis (Arezi and Hogrefe, 2009; Baranauskas *et al*., 2012; Matamoros *et al*., 2013; Mohr *et al*., 2013). Improvements in the intrinsic fidelity of RTs may have a positive impact on the reliability of whole transcriptome shotgun sequencing (i.e., RNA-seq) data (Ozsolak and Milos, 2011).

1.4.3.1. Experimental methods to estimate the accuracy of RTs *in vitro*

1.4.3.1.1. Biochemical assays

The fidelity of purified RTs has been analyzed *in vitro* by using gel-based assays. The fixation of a base substitution involves the incorporation of an incorrect nucleotide, followed by the extension of the mismatched primer. These events can be qualitatively detected by dNTP exclusion assays, and quantitatively analyzed by enzymological tests involving the determination of enzyme kinetics for the incorporation of correct and incorrect nucleotides.

The dNTP exclusion assays (also refer to as "minus"-sequencing gel assays) provide a rough estimate of fidelity based on the observed primer extension efficiencies (Sanger and Coulson, 1975; Preston *et al*., 1988). Three dNTPs, each one at a relatively high concentration, are used to assess misincorporation opposite multiple template sites. Faithful enzymes are expected to stop synthesis in the position where the absent dNTP should be incorporated. Thus, full-extension bands require insertion of incorrect nucleotides beyond this position. The analysis of the pattern of bands generated by each RT gives an idea of its fidelity.

In nucleotide misinsertion assays, a binary complex RT/template-primer is formed (Preston *et al*., 1988), and the ability to extend the 3' end of the primer is determined in the presence of increasing concentrations of correct or incorrect nucleotides at different times. The extended primer is measured by quantitative gel electrophoresis; and then kinetic parameters showing the efficiency of nucleotide incorporation are obtained by fitting the resulting data to the Michaelis-Menten equation. The turnover number (k_{cat}) and the Michaelis constant for dNTP (K_m) can be determined under steady-state conditions. However, these parameters are not appropriate to obtain mechanistic insight into the DNA polymerization reaction, since their values are strongly influenced by the slow rate of dissociation of the enzyme from the template-primer (k_{off}) (Kati *et al*., 1992).

Relevant parameters in these assays are obtained under rapid transient kinetics methods, that allow the measurement of the equilibrium dissociation constant (K_d) for the interaction of dNTP with the binary complex RT/template-primer, as well as the nucleotide incorporation rate (k_{pol}) (**Figure 6**). These kinetics are determined under pre-steady-state conditions, after time intervals ranging from 3 ms to several seconds (reviewed in Menéndez-Arias, 2002, 2009). The catalytic efficiency of the reaction can be determined by the ratio k_{pol}/K_d (or k_{cat}/K_m). Thus, misinsertion efficiencies (f_{ins}) are defined as the catalytic efficiency of incorporation of each incorrect nucleotide relative to the correct one on a specific template-primer, and reflect the nucleotide selectivity of the enzyme (reviewed in Echols and Goodman, 1991; Johnson, 1993):

$$f_{\mathrm{ins}} = \frac{(k_{\mathrm{pol}}/K_{\mathrm{d}})\ \text{incorrect nucleotide}}{(k_{\mathrm{pol}}/K_{\mathrm{d}})\ \text{correct nucleotide}}$$

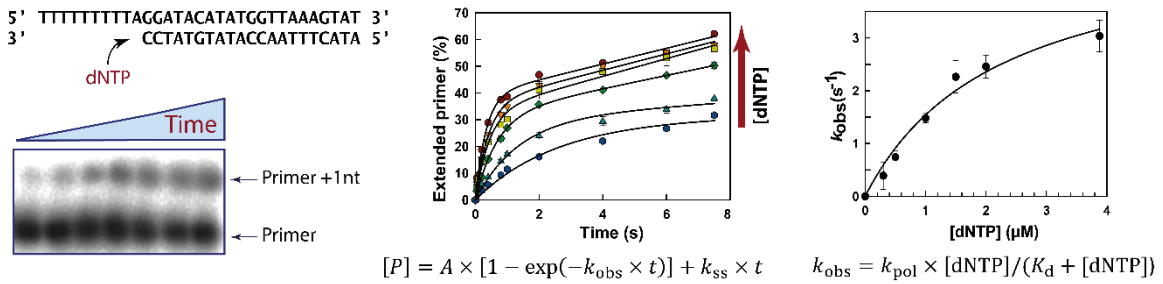


Figure 6. Determination of pre-steady-state kinetic constants through one-nucleotide incorporation assays under heteropolymeric template-primers. The proportion of extended primer at position +1 is quantified by densitometry at different times and dNTP concentrations. Results are fitted to a first order equation (below the center panel), where A is the amplitude, *t* is the reaction time, k_{obs} is the apparent kinetic constant of formation of the phosphodiester bond, and k_{ss} is the kinetic constant of the steady-state linear phase. The dependence of k_{obs} on the dNTP concentration is described by the hyperbolic equation of Michaelis-Menten, where K_d and k_{pol} are the catalytic and affinity constants for the dNTP incorporation, respectively. Adapted from Barrioluengo (2013).

The rate of polymerization (k_{pol} or k_{cat}) is expected to be higher for the correct nucleotide than for the incorrect ones. In contrast, incorrect nucleotides bind the RT/template-primer complex with less affinity, and show a higher K_d or K_m value. Therefore, higher fidelity implies lower misinsertion efficiency. The method is also adequate to compare the incorporation of dNTPs *versus* rNTPs, dNTPs *versus* ddNTPs, or natural *versus* nucleotide analogues (reviewed in Menéndez-Arias, 2002).

Mispair extension assays involve the determination of kinetic constants (k_{pol} and K_d) for the incorporation of a correct nucleotide, using template-primer complexes with either matched or mismatched 3' primer termini. The efficient extension of mismatched complexes has been recognized as a major determinant of the low accuracy of retroviral RTs (Perrino *et al*., 1989; Yu and Goodman, 1992). In these assays, mispair extension ratios (f_{ext}) can be obtained as follows:

$$f_{ext} = \frac{(k_{pol}/K_d)\ \text{mismatched primer}}{(k_{pol}/K_d)\ \text{matched primer}}$$

Biochemical assays allow a proper comparison of the mechanistic insights into the DNA synthesis activity of multiple RTs. However, these assays have several limitations. First, reaction conditions such as pH, dNTP concentrations, the nature and concentration of the divalent cation (e.g., Mg^{+2}, Mn^{2+}), and the sequence of the template-primer can all significantly impact the observed selectivity rate (Brosius *et al*., 1983; Beckman *et al*., 1985; Eckert and Kunkel, 1990, 1993a, 1993b; Bebenek *et al*., 1992; Bakhanashvili and Hizi, 1992a, 1992b, 1993; Achuthan *et al*., 2014). Second, the stability and structure of the template-primer complex and its ratio to the RT may impact the observed results. Third, gel-based experiments are restricted to the analysis of a small number of incorporation sites, with only one or a few target nucleotides, which may not represent the overall error rate (reviewed in Svarovskaia *et al*., 2003).

1.4.3.1.2. Genetic assays

Many genetic assays have been used to estimate the fidelity of numerous polymerases by measuring mutant frequencies in a reporter gene. These assays can be classified into two major groups, including those designed to detect the inactivation of a reporter gene (forward mutation assays) or the reversion of an inactivating mutation (codon reversion assays). The reporter genes should be easily selected or phenotypically identifiable (e.g., *lac*Z gene or its truncated peptide *lacZ*α).

Codon reversion assays were initially used to determine base substitution rates of DNA polymerases (Weymouth and Loeb, 1978). In these assays, the bacteriophage ssDNA that was used as template for DNA synthesis contained an in-frame termination codon in essential phage genes. Typically, it is an amber (TAG) mutant in the phage ΦX174 DNA (Preston *et al*., 1988; Weber and Grosse, 1989), an ochre (TAA) mutant in phage T4 DNA (Orias and Gartner, 1966; Soll and Berg, 1969; Ohta *et al*., 2002), or an opal (TGA) mutant at position 89 of the *lacZ* gene in phage M13mp2 DNA (Roberts *et al*., 1988; Kunkel and Soni, 1988). The reversion of nonsense codons to a coding triplet can be visualized as a revertant plaque upon transformation of host cell bacteria with copied DNA or as a dark blue M13 plaque using the appropriate conditions (Preston *et al*., 1988; Roberts *et al*., 1989). However, these methods only focus on one or a few sites, and estimates are restricted to specific base substitution errors (mostly, G-to-A). Because retroviral mutation rates are highly sequence-dependent, the target nucleotides chosen in a reversion assay probably do not reflect the overall mutation rate.

In contrast, forward mutation assays score for nucleotide changes at many different sites and in a wide variety of sequence contexts, thereby providing information not only on base substitutions, but also on frameshifts. Besides, these assays may help the identification of relevant hotspots. Several reporter genes used in forward mutation assays are HIV-1 *env* gene variable region 1 (V1) (cloned in M13) (Ji and Loeb, 1994), and the truncated *lacZ*α gene (cloned in vectors pBluescript or Litmus 29, and in phage M13mp2) (Ji and Loeb, 1992; Bebenek and Kunkel, 1995; Boyer and Hughes, 2000).

Using this method, some mutations in the reporter gene remain undetectable by not producing a noticeable phenotypic change. Detailed information based on experimental data of mutational target sites that reliably result in a detectable mutant phenotype is available for *lacZ*α gene, but not for the complete *lac*Z gene or other reporters (Svarovskaia *et al*., 2003). Then, *lacZ*α-based forward mutation assays do not depend on assumptions of which mutations lead to a detectable phenotypic change, being the most widely used to measure the fidelity of purified RTs (**Table 1**). In these assays, a gapped DNA from phage M13mp2 (dsDNA lacking the *lacZ*α gene in one of the two strands) is used as substrate of DNA synthesis reactions catalyzed by the RT. Errors produced by the enzyme

while copying the gapped region may generate β-galactosidase enzymes with lower activity, and are detected as colorless or light blue plaques after transformation in appropriate bacterial strains, in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and isopropyl β-D-1-thiogalactopyranoside (IPTG) (Roberts *et al*., 1988).

Table 1. Estimated error rates of retroviral RTs in comparison with wild-type HIV-1 RT (group M subtype B) determined with M13mp2-based forward mutation assays using *lacZα* as reporter gene

RT	Mutation rate	Accuracy [a] (fold change)	HIV-1 RT	References
HIV-1 group O	5.5×10^{-5}	2.5	BH10 (p66/p51)	Álvarez *et al*., 2009
SIV_{agm}	2.9×10^{-5}	1.8	HXB2 (p66/p51)	Stuke *et al*., 1997
SIV_{mne}	1.6×10^{-5}	9.7	BH10 (p66/p51)[b]	Diamond *et al*., 2003
FIV	6.2×10^{-5}	2.8	BH10 (p66/p51)[b]	Operario *et al*., 2005
MLV	3.3×10^{-5}	14.3	(p66/p61)[c]	Roberts *et al*., 1988, 1989
	1.2×10^{-5}	11.7	BH10 (p66/p51)	Barrioluengo *et al*., 2012
XMRV	1.2×10^{-5}	11.7	BH10 (p66/p51)	Barrioluengo *et al*., 2012
FeLV	5.8×10^{-6}	29	BH10 (p66/p51)[b]	Operario *et al*., 2005
AMV	5.9×10^{-5}	9.5	(p66/p61)[c]	Roberts *et al*., 1988, 1989
PFV	1.7×10^{-4}	0.4	BH10 (p66/p51)	Boyer *et al*., 2007

SIV, simian immunodeficiency virus (agm, African green monkey; mne, pigtail macaque); FIV, feline immunodeficiency virus; MLV, murine leukemia virus; XMRV, xenotropic murine leukemia virus-related virus, FeLV, feline leukemia virus; AMV, avian myeloblastosis virus; PFV, prototype primate foamy virus.
[a] Increase relative to the WT HIV-1 (group M subtype B) RT of each experiment. The error rates for enzymes of subtype B were in the range of $0.6\text{-}6.7\times10^{-4}$ (reviewed in Menéndez-Arias, 2009). [b] Data for HIV-1 RT were not provided in the referenced papers. For comparative purposes, mutation rates for HIV-1 RT obtained in the same lab were taken from Weiss *et al*., 2004. [c] Purified from HIV-1 group M subtype B virions (unreported strain) (Roberts *et al*., 1988).

1.4.3.1.3. Deep sequencing assays

Deep sequencing technologies are commonly used for extensive sampling of viral populations and cellular transcriptomes (Capobianchi *et al*., 2013; Brancaccio *et al*., 2018; Gallego *et al*., 2018; Pennington *et al*., 2018). However, protocols can be adapted for the analysis of RT's fidelity while synthesizing cDNA in reactions carried out *in vitro* (Gordon *et al*., 2015; de Paz *et al*., 2018; Hebert *et al*., 2018). These outbreaking methodologies yield thousands of nucleotides, which may facilitate the calculation of error rates of high-fidelity enzymes. The main advantage of this approach is that all types of errors can be detected regardless of whether they give rise to a phenotypic change or not, and the analysis can be performed in hundreds of sites with a wide range of sequence contexts. However, deep sequencing also presents several restrictions that bias the sampling. Among them, errors introduced during the polymerase-chain reaction (PCR) steps and throughout the sequencing protocol are probably the most important ones in inflating the heterogeneity of the sample. In addition, differential amplification can skew allele/mutations frequency, and PCR-mediated recombination can disrupt linkage (Jabara *et al*., 2011).

Many different protocols have been designed to address these issues. Barcoding of nucleic acid molecules before sequencing has become one of the most recurrent solutions, since it could facilitate the discrimination of sequencing errors from genuine mutations, by creating a consensus between all reads with the same barcode. Barcodes were applied in different steps of reverse transcription and sequencing processes, each one allowing the discrimination or elucidation of a specific source of errors (**Figure 7**). First, transcription errors can be elucidated by attaching the barcode at the 5' end of the RNA molecules used as template, as in the Replicated sequencing (Rep-Seq) protocol (Gout *et al*., 2013).

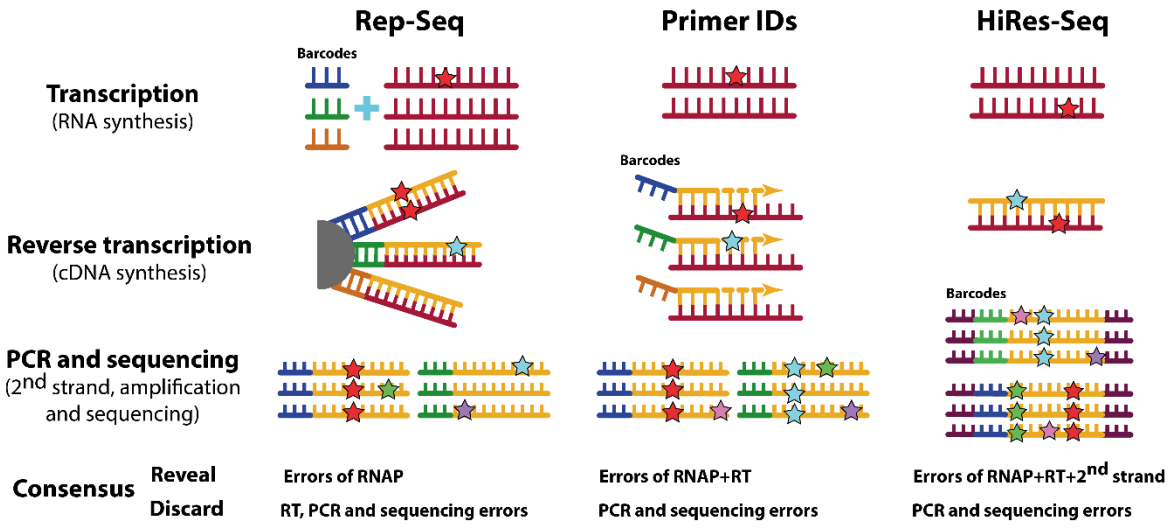


Figure 7. Summary of several next-generation sequencing technologies adapted to study polymerase fidelity by using barcodes before sequencing. Replicated sequencing (Rep-Seq) (Gout *et al*. 2013) identifies transcription errors (red stars) by attaching individual barcoded mRNAs to beads and reverse transcribing them multiple times. Mutational changes in common to all cDNAs containing the same barcode represent transcription errors. Additional errors are shown as stars of different colors: blue, for those made during reverse transcription; green, for errors made during the synthesis of the second strand (i.e., completion of dsDNA using the cDNA obtained in the reverse transcription as template); purple and pink, for amplification and sequencing errors. This criteria is maintained in the whole picture. Primer IDs (Jabara *et al*., 2011) barcode each cDNA molecule, revealing the combination of transcription and reverse transcription errors. High-resolution RNA sequencing (HiRes-Seq) (Imashimizu *et al*., 2013) include barcodes in the PCR primers, discarding amplification and sequencing errors. RNAP, RNA polymerase.

Second, the barcodes can be introduced in the primer used for reverse transcription (also known as Primer IDs) (Jabara *et al*., 2011; Zhou *et al*., 2015), discriminating the PCR and sequencing errors, and obtaining a combined error rate between transcription and RT-mediated RNA-dependent DNA synthesis. Moreover, amplicons or cDNA molecules can be tagged at their termini by ligation of adaptors harboring random nucleotide sequences (Kinde *et al*., 2011; Schmitt *et al*., 2012; Kennedy *et al*., 2014). Finally, in the first PCR, including the barcode in the primers used for amplification may enable to discount PCR and sequencing errors during analysis of the high-resolution RNA-sequencing method (HiRes-Seq) (Imashimizu *et al*., 2013).

Of all these methods, the most convenient to measure the fidelity of RTs is that of PrimerIDs, that allows to tag each cDNA molecule giving them a unique identity. However, the diversity of techniques arises from the complexity of the pipelines used for producing libraries and analyzing results by bioinformatics. A standardize and consistent measurement of RT's fidelity by next-generation sequencing platforms is still missing.

1.4.3.2. DNA-dependent DNA synthesis fidelity of WT and multidrug-resistant HIV-2 RTs

Fidelity of WT and mutant HIV-1 RTs has been extensively studied (for a review see Menéndez-Arias, 2002, 2009, 2013a; Svarovskaia *et al*., 2003). However, attempts to correlate mutation frequencies in HIV-1 and HIV-2 with the attenuated progression to AIDS have been inconclusive (Lemey *et al*., 2007; MacNeil *et al*., 2007; Skar *et al*., 2010). Several factors including dNTP pools, viral and cellular proteins and the cell types used in the assays may have an important effect on the viral mutation rate. Next-generation sequencing analysis of amplicons obtained from infected cells showed that HIV-2 had lower mutation rates than HIV-1, although the observed differences were due to the high frequency of G-to-A transitions found in HIV-1 (Rawson *et al*., 2015). The higher susceptibility of this virus to APOBEC3-mediated hypermutation could explain these results (Ribeiro *et al*., 2005). Nucleotide incorporation assays carried out with heteropolymeric template-primers under steady-state conditions revealed only subtle differences in misinsertion and mispair extension fidelity between the prototypic HIV-1_{BH10} and HIV-2_{ROD} RTs (Bakhanashvili and Hizi, 1993). However, proper assessments of the intrinsic fidelity of the HIV-2 RT, including estimates of base substitution and frameshift error rates relative to other RTs have not been reported.

HIV-2 shows natural resistance to a number of antiretroviral drugs designed to suppress HIV-1 propagation. Thus, NNRTIs (**Figure 8A**), the fusion inhibitor enfuvirtide and several protease inhibitors are ineffective against HIV-2, while the clinical efficacy of maraviroc is uncertain (reviewed in Menéndez-Arias and Álvarez, 2014). Unlike HIV-1, HIV-2 does not develop resistance to nucleoside RT inhibitors (NRTIs) (**Figure 8B**) through the excision pathway (involving amino acid substitutions M41L, D67N, K70R, L210W, T215F/Y and K219E/Q in the viral HIV-1 RT (reviewed in Menéndez-Arias, 2013b)), but relies exclusively on nucleotide discrimination (Boyer *et al*., 2012).

Clinical and epidemiological studies have shown that HIV-2 has lower genetic barriers than HIV-1 to the development of multidrug resistance (Gottlieb *et al*., 2009). NRTIs (usually tenofovir/emtricitabine) constitute the backbone of commonly prescribed therapies against HIV-2 (Menéndez-Arias and Álvarez, 2014; Günthard *et al*., 2016). RT substitutions K65R and Q151M are frequently identified in virus obtained from treated HIV-2-infected patients (Ntemgwa *et al*.,

2009), and the presence of both changes together with M184V confers classwide NRTI resistance (Smith *et al*., 2009). An estimated prevalence of 9% has been reported for the combination of K65R, Q151M and M184V in a European cohort of HIV-2-infected patients treated with antiretroviral drugs (Deuzing *et al*., 2015).

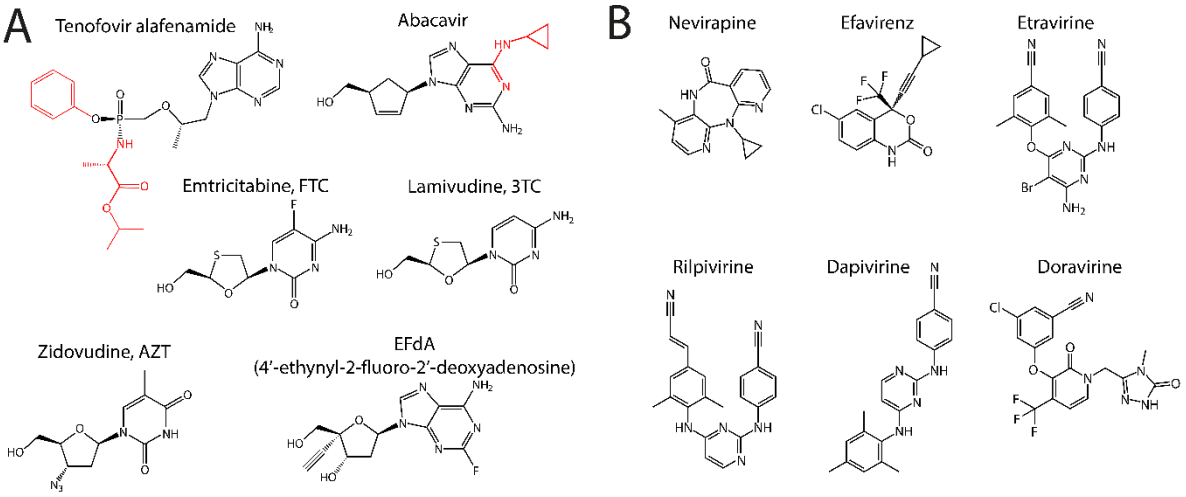


Figure 8. Chemical structures of antiretroviral drugs active against HIV-1 RT. (A) Chemical structures of nucleoside analogues RT inhibitors (NRTIs). Tenofovir alafenamide, emtricitabine, abacavir, lamivudine and zidovudine are approved drugs commonly used to treat HIV-infected patients. Tenofovir alafenamide is a prodrug currently preferred over tenofovir disoproxil fumarate due to its reduced renal and bone toxicity (Ray *et al*., 2016). EFdA is a nucleoside analogue in preclinical development. Atoms modified during the conversion of the analogues to metabolically relevant derivatives are indicated in red. **(B)** Chemical structures of nonnucleoside RT inhibitors. Adapted from Menéndez-Arias *et al*. (2017).

Phenotypic studies using recombinant HIV-2 have shown that Q151M alone confers high-level resistance to zidovudine (AZT), and moderate- to low-level resistance to stavudine, abacavir, didanosine and emtricitabine. However, it has a small impact on the viral susceptibility to lamivudine and tenofovir (Damond *et al*., 2005; Smith *et al*, 2009; Deuzing *et al*., 2015). M184V confers high-level resistance to approved cytidine analogues (i.e., lamivudine and emtricitabine) (Smith *et al*., 2009), while K65R confers low-level resistance to didanosine and moderate to high-level resistance to cytidine analogues in HIV-2 strains (Damond *et al*., 2005; Smith *et al*., 2009). As previously shown in HIV-1 (Menéndez-Arias, 2013b), K65R arises in HIV-2-infected patients exposed to tenofovir and other NRTIs (Ntemgwa *et al*., 2009; Peterson *et al*., 2011). In addition, K65R can be selected *in vitro* after HIV-2 exposure to increasing doses of tenofovir (Andreatta *et al*., 2013), and produces a 2- to 7-fold decrease in viral susceptibility to the drug (Andreatta *et al*., 2013; Deuzing *et al*., 2015).

Lys65, Gln151 and Met184 locate at the dNTP binding site of the RT (Huang *et al*., 1998) (**Figure 4B**). By themselves, amino acid substitutions K65R, Q151M or M184V in HIV-1 RT confer reduced catalytic efficiency of nucleotide analogue incorporation (reviewed in Menéndez-Arias,

2013b). Consistently with those findings, mutant HIV-1 with RT substitutions K65R or M184V show delayed replication kinetics in comparison with the WT virus (Back *et al*., 1996; Cong *et al*., 2007). In the case of M184V, the fitness defect is more pronounced at low dNTP concentrations and is expected to be more relevant for viruses replicating in non-dividing cells (e.g., macrophages) (Keulen *et al*., 1997). HIV-1 RTs with Arg65 instead of Lys showed decreased nucleotide incorporation rates (k_{pol}) in pre-steady-state kinetic assays (Deval *et al*., 2004, 2005; Barrioluengo *et al*., 2011). Other studies have shown that K65R produces a > 8-fold increase in the intrinsic fidelity of the HIV-1 RT, an effect that has been demonstrated with RTs of highly divergent HIV-1 group M (subtype B) and group O strains (Shah *et al*., 2000; Barrioluengo *et al*., 2011). In agreement with enzymatic data, K65R produced a 3.3-fold mutant frequency reduction in single round of replication assays, using an HIV-1_{NL43} vector containing the *lacZ* gene (Mansky *et al*., 2003).

In contrast to the well-studied HIV-1 RT, studies on the effects of NRTI resistance-associated mutations on the nucleotide specificity and fidelity of the HIV-2 polymerase are limited to a few single- and double-mutants with substitutions at positions 74, 89, 111 and 215 (i.e., L74V, E89G, V111I, S215Y, L74V/S215Y and E89G/S215Y), whose effects were evaluated for a small subset of misincorporations or mispairs using nucleotide discrimination assays (Taube *et al*., 1997; Deuzing *et al*., 2015). Unfortunately, those studies could not provide reliable mechanistic interpretations of the data because assays were performed under steady-state conditions. None of the studied residues was part of the complex conferring multi-NRTI resistance, although sometimes V111I appears as an accessory mutation that compensates for the lower fitness of classwide NRTI-resistant HIV-2 (Deuzing *et al*., 2015). Furthermore, as for the WT HIV-2 RT, it has not been tested the intrinsic fidelity of mutant HIV-2 RTs carrying amino acid substitutions associated to multidrug resistance.

1.4.3.3. RNA-dependent DNA synthesis fidelity of retroviral RTs

Despite the large amount of research on the fidelity of retroviral RTs, most of the available studies have been devoted to the analysis of DNA-dependent DNA synthesis accuracy (reviewed in Menéndez-Arias, 2002, 2013a). These studies have shown that oncoretroviral RTs are more faithful than lentiviral RTs. Thus, the intrinsic error rates of HIV-1 RT were more than 10-fold higher than those obtained with AMV and MLV RTs (Roberts *et al*., 1988, 1989; Barrioluengo *et al*., 2012), determined with M13mp2 *lacZα* forward mutation assays (Bebenek and Kunkel, 1995). However, in HIV-1 RTs, antiretroviral drug resistance-associated mutations such as K65R or the combination of K65R and V75I were shown to increase fidelity of DNA-dependent DNA synthesis to levels similar to those obtained with oncoretroviral RTs (Shah *et al*., 2000; Barrioluengo *et al*., 2011).

The fidelity of RNA-dependent DNA synthesis of retroviral RTs has been evaluated in enzymatic assays by comparing kinetic parameters for misincorporation and mispair extension. Assays carried out under steady-state conditions with AMV, MLV and HIV-1 RTs did not reveal large differences when DNA templates were substituted by RNA templates (Yu and Goodman, 1992; Bakhanashvili and Hizi, 1993). However, using pre-steady-state kinetics, Kerr and Anderson (1997) showed that misinsertion fidelity was 9-64 times (with duplex 45/25mer) and 14-23 times (with duplex 45/22mer) higher in RNA-templated than in DNA-templated reactions catalyzed by HIV-1 RT. These experiments were performed with synthetic duplexes bearing the same nucleotide sequences (except for having U instead of T in the RNA templates). However, nucleotide incorporation assays are restricted to a few template-primers and therefore provide limited information on the propensity of RTs to introduce base substitutions.

Attempts to compare the fidelity of retroviral RTs in DNA polymerization reactions carried out with RNA or DNA templates using genetic assays did not provide consistent results. Studies performed with HIV-1 RT and the amber16 reversion assay using phage ΦX174 showed that two out of the seven specific mismatches analyzed had 20- and 7-fold lower mutation frequencies with DNA templates than with RNA templates (Hübner *et al*., 1992). On the other hand, M13-based assays using *lacZα* or the HIV-1 *env* hypervariable region 1 (V1) as target genes showed little differences in fidelity between reactions carried out with HIV-1 RT using RNA or DNA templates. Thus, base substitution error rates determined with *lacZα* were 1.7×10^{-4} and 1.4×10^{-4} for DNA-dependent and RNA-dependent reactions, respectively (Ji and Loeb, 1992). These small differences were also found for overall error rates in assays using a fragment of *env* as a target sequence (i.e., 1.9×10^{-4} with DNA templates and 2.0×10^{-4} with RNA templates) (Ji and Loeb, 1994). In contrast, and using a modified version of the M13mp2 *lacZα* forward mutation assay, Boyer *et al*. (1992) showed that the overall fidelity of HIV-1 RT was about two to six-fold higher while copying RNA templates than DNA templates. In these experiments, authors suggested the contribution of errors made by the T7 RNAP while preparing the RNA used as template in the fidelity assays, acknowledging an undetermined impact on the calculated error rates.

The discrepancies in the fidelity assessments could be explained in part by methodological differences in the assays, such as the use of distinct templates and RTs, or the different WT HIV-1 group M/subtype B RT used for comparison in these experiments. For example, homodimers (p66/p66) *versus* heterodimers (p66/p51), or RT variants derived from different viral strains, such as NL4-3, HXB2, BH10 or NY5. Thus, the intrinsic fidelity of retroviral RTs during RNA-dependent DNA synthesis reactions should be properly measured under the same conditions, yielding a complete analysis of mutational spectra and estimates of base substitution and frameshift error rates.

2. OBJECTIVES

- **Objective 1.** To determine the DNA-dependent DNA synthesis accuracy of WT HIV-2_{ROD} RT, as well as that of mutant RTs carrying amino acids substitutions associated to multidrug-resistance such as K65R, and the combination K65R, Q151M and M184V by using M13mp2 forward mutation assays based on the expression of *lacZα*.

- **Objective 2.** To characterize the RNA-dependent DNA synthesis fidelity of retroviral RTs, including those of WT MLV, AMV, HIV-1_{BH10}, HIV-1_{ESP49}, as well as mutant HIV-1_{ESP49} RTs carrying the amino acid substitutions K65R/V75I, by using forward mutation assays.

- **Objective 3.** To determine mutant frequencies, error rates for different type of mutations (e.g., base substitutions, frameshifts) and mutational spectra of retroviral RTs while reverse transcribing RNA templates, and to compare them with those obtained in DNA templates.

- **Objective 4.** To study the impact of heterogeneity present in RNAs used as templates of reverse transcription in the estimates of RT's fidelity.

- **Objective 5.** To optimize deep sequencing experiments based on the use of Primer IDs to analyze the fidelity of RTs on RNA templates.

3. MATERIALS AND METHODS

3.1. Enzymes and bacterial strains

WT and mutant RTs belonging to different lentiviruses (i.e., HIV-1 and HIV-2) and oncoretroviruses (i.e., AMV and MLV) were used in this work (**Table 2**). *Escherichia coli* strains (**Table 3**) and plasmids (**Table 4**) used in this Thesis are indicated, together with their main features.

Table 2. RTs and RNAPs used in this study

Enzymes	GenBank ID/ no. catalogue	References and companies
Recombinant		
RTs		
HIV-1_{BH10}	AAA44198.1[a]	Ratner *et al*. (1985)
HIV-1_{ESP49}	AAC64709.1[b]	Quiñones-Mateu *et al*. (1998)
$K65R_{ESP49}$		
$K65R/V75I_{ESP49}$		
HIV-2_{ROD}	X05291.1[c]	Clavel *et al*. (1986); Guyader *et al*. (1987)
$K65R_{ROD}$		
$K65R/Q151M/M184V_{ROD}$		
RNAPs		
Bacteriophage T7	AAA32569.1[d]	Davanloo *et al*. (1984)
Commercial		
RTs		
AMV	#M170A	Promega
MLV	#M5101	Promega
RNAPs		
Bacteriophage T7	#P2075	Promega
	#AM1312	Ambion

HIV-1_{BH10}, WT HIV-1 group M subtype B strain BH10; HIV-1_{ESP49}, WT HIV-1 group O strain ESP49; $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$, RTs of HIV-1_{ESP49} bearing amino acid substitutions K65R and K65R/V75I, respectively; HIV-2_{ROD}, WT HIV-2 group A strain ROD; $K65R_{ROD}$ and $K65R/Q151M/M184V_{ROD}$, RTs of HIV-2_{ROD} bearing the change K65R or the combination K65R, Q151M and M184, respectively.
[a] The amino acid sequence of HIV-1_{BH10} RT used in this Thesis has a 100% identity with the positions 168−560 of the referred Pol polyprotein. [b] The amino acid sequence of HIV-1_{ESP49} RT presents changes V142I, S204L, V314M, and R398W in comparison with the given GenBank ID. [c] The amino acid sequence of HIV-2_{ROD} RT is 100% identical to the translated nucleotides of the HIV-2_{ROD} RNA genome from position 2,381 to 4,057 of the GenBank ID reference. [d] The T7 RNAP contains changes R389K, H623Y, and P665L compared to the provided GenBank ID.

Table 3. Bacterial strains and plasmids referred to in this study

Designation	Genotype or description
***E. coli* strains**	
XL1 Blue	*recA*1, *endA*1, *gyrA*96 (nal^R), *thi*-1, *hsdR*17 (r_K^- m_K^+), *glnV*44, *relA*1, *lac*, F' [Tn10 $proA^+B^+$ $lacI^q$ Δ(lacZ)M15]
DH5α	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG phoA purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-
BL21	B, F^- *omp*T *gal dcm lon hsd*S_B(r_B^- m_B^-)
BL21(DE3)pLysS	BL21 strain, λ(DE3 [*lac*I *lac*UV5-T7 gene 1 *ind1 sam7 nin5*]), with pLysS
BL21(pREP4)	BL21 strain, with plasmid pREP4
NR9099	[Δ(*pro-lac*)], *thi, ara, recA56/F'* (*proAB, lacIqZΔM15*)
MC1061	*hsdR, hsdM+, araD,* Δ(*ara, leu*), Δ(*ara, leu*), Δ(*laclPOZY*), *galU, galK, strA*
CSH50	[Δ(*pro-lac*)], *thi, ara, strA/F'* (*proAB, laclqZΔM15, traD36*)

Table 4. Bacterial strains and plasmids referred to in this study

Designation	Genotype or description
Plasmids	
$pRT66B_{BH10}$	trcP vector, *lacI*^{q}, Amp^{R}, MscI-HindIII fragment, encodes the HIV-1_{BH10} p66 subunit
$pRT66B_{ESP49}$	trcP vector, *lacI*^{q}, Amp^{R}, MscI-HindIII fragment, encodes the HIV-1_{ESP49} p66 subunit
pATPR	tacP vector, Tn3R, Kan^{R}, EcoRI-BamHI fragment, encodes the HIV-1_{817} PR
pT5m	T7P vector, Amp^{R}, BamHI-HindIII fragment, encodes the HIV-2_{ROD} p68 subunit
pLysS	[*T7p20 ori*$_{p15A}$] (Cm^{R}) ; encodes the T7 lysozyme
pQE9T7	T5P vector, Amp^{R}, Cm^{R}, BamHI-HindIII fragment, encodes the T7 RNAP
pREP4	[*ori*$_{p15A}$] *lacI*^{q}, Kan^{R}; encodes the *lac* represor

3.2. RT purification

3.2.1. HIV-1 RTs

HIV-1 RT purification relies on the co-expression of the large subunit of the RT (p66) and the HIV-1 protease in *E. coli* XL1-Blue (**Table 3**), previously transformed with plasmids $pRT66B_{BH10}$ or $pRT66B_{ESP49}$, together with pATPR (**Figure 9**) (Boretto *et al*., 2001; Matamoros *et al*., 2005).

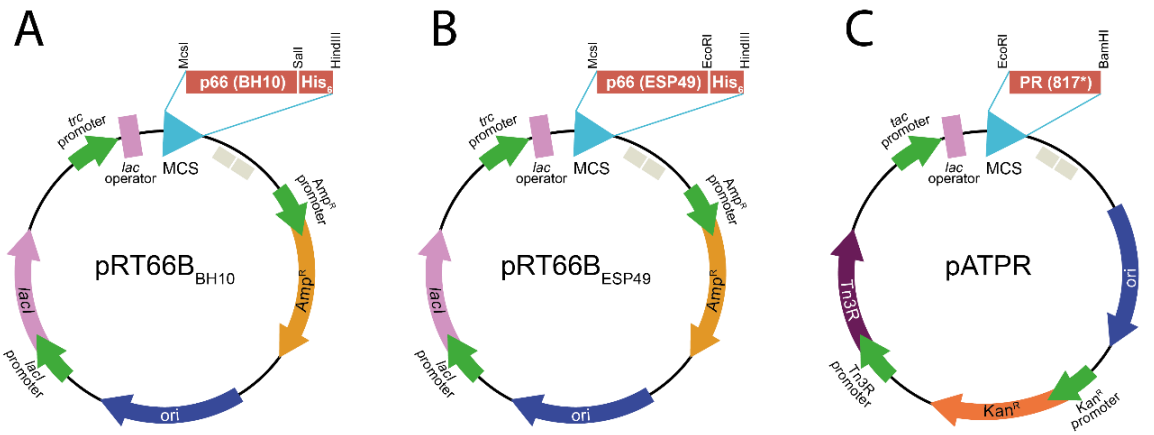


Figure 9. Plasmids used in HIV-1 RTs purification. (A) Plasmid $pRT66B_{BH10}$. This construction derives from plasmid pTrc99A (Amersham Biosciences) (Canard *et al*., 1999). The plasmid $pRT66B_{BH10}$ encodes the major subunit (p66) of the RT from HIV-1 group M subtype B strain BH10, including a region with a tag of six histidines at the C-terminus of the enzyme. The insert was introduced using the restriction sites of MscI and HindIII endonucleases in the multicloning site (MCS) (Boretto *et al*., 2001). **(B)** Plasmid $pRT66B_{ESP49}$. Same as the plasmid represented in A, but codifying for the p66 subunit of the HIV-1 group O strain ESP49 RT (Matamoros *et al*., 2005; Álvarez *et al*., 2009). **(C)** Plasmid pATPR. This plasmid contains the HIV-1 protease coding sequence preceded by two ribosome binding sites (RBS), including that of T7g10 before the starting ATG codon of the viral protease. The insert was cloned into the EcoRI and BamHI sites of the plasmid pAT, from which pATPR proceeds (Boretto *et al*., 2001). The co-expression of pATPR together with either $pRT66B_{BH10}$ or $pRT66B_{ESP49}$ produces p66/p51 heterodimers resulting from cleavage of p66 by the viral protease. Amp^{R} and Kan^{R} indicate the regions codifying for the ampicillin and kanamycin resistance genes, respectively. His_6 represents the sequence that encodes the six-histidine tag. Tn3R, Tn3 resolvase; ori, origin of replication. The gen *lac*I encodes the lac repressor protein that binds to the *lac* operator and allows to control the expression of HIV-1 p66 and protease polypeptides by addition of IPTG.

3.2.1.1. Site-directed mutagenesis

Nucleotide sequences encoding for p66 subunits of HIV-1_{BH10} and HIV-1_{ESP49} RTs were cloned in plasmids pRT66B_{BH10} and pRT66B_{ESP49}, respectively (**Figure 9**). The vector pRT66B_{ESP49} was used as template to introduce mutations K65R and K65R/V75I using the "QuickChange™ Site-Directed Mutagenesis Kit" of Stratagene according to the manufacturer's instructions. Mutagenic primers are shown in **Table 5**.

Table 5. Primers used in mutagenesis reactions with pRT66B_{ESP49}

Mutation	Mutagenic primer sequence
K65R	5′-CTTTGCTATAAAAAGGAAAGATAGTACTAAGTGG-3′ 3′-GAAACGATATTTTTCCTTTCTATCATGATTCACC-5′
V75I	5′-CAGTACTAAATGGAGAAAATTAATAGATTTCAGAGA-3′ 3′-GTCATGATTTACCTCTTTTAATTATCTAAAGTCTCT-5′

Left column indicates the mutation introduced in the RT. Mutagenized codons are underlined.

Mutagenesis reactions were carried out in 50 µl of 20 mM Tris-HCl pH 8.8 buffer, containing 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% (w/v) Triton X-100, 0.1 mg/ml bovine serum albumin (BSA), 125 ng of each of the corresponding mutagenic primers, around 50 ng of the template plasmid, 500 µM of each of the four dNTPs and 2.5 U of Pfu Turbo DNA polymerase (Stratagene). The PCR was started with an incubation of 2 min at 95 ºC followed by 20 cycles each one including incubations of 30 s at 95 ºC, 1 min at 55 ºC and 12 min at 68 ºC. After cooling the reaction at 4 ºC, the obtained product was treated with 10 U of DpnI (New England Biolabs) for 1 hour at 37 ºC. DpnI is an endonuclease that specifically degrades the methylated or hemimethylated DNA. The template pRT66B_{ESP49} used in the mutagenic reaction is methylated due to its amplification in an *E. coli dam*$^+$ strain. After that treatment, the newly synthesized DNA with the desired mutations remains undigested, and was analyzed on a 1% agarose gel.

E. coli DH5α competent cells (**Table 3**) were then transformed with 5 µg of the product of the mutagenesis reaction by thermal shock (20-30 min on ice, followed by 1 min at 42 ºC and 2 min on ice). Then, bacteria was grown on 1 ml of Luria-Bertani medium (LB, containing 10 g bactotriptone, 5 g yeast extract and 10 g NaCl, per liter) during 1 hour at 37 ºC, and ampicillin-resistant colonies were selected. The integrity of the entire RT sequence, as well as the presence of the desired mutations, were confirmed by sequencing the DNA plasmid extracted with the "Wizard Plus SV Minipreps" kit (Promega). The sequencing reactions were carried out in Macrogen Korea (Seoul, South Korea), and sequences were subsequently analyzed using Seqman 4.0 (DNAstar Inc.) software. Once the presence of the mutation was confirmed, the plasmids were used to transform the *E. coli* XL1-Blue PR strain by thermal shock. This strain derives from *E. coli* XL1-Blue, previously transformed with the pATPR plasmid (**Figure 9**). This plasmid contains the nucleotide sequence encoding the HIV-1 protease from isolate 817 (GenBank ID: AF117566), except for a

V64I substitution (Boretto *et al*., 2001), together with the kanamycin resistance gene. Ampicillin- and kanamycin-resistant colonies were selected for the following steps.

3.2.1.2. Expression of HIV-1 RT p66/p51 heterodimers

Twenty ml of LB medium containing ampicillin (100 μg/ml) and kanamycin (50 μg/ml) were inoculated with 5-10 μl of a glycerol-containing stock of *E. coli* XL1-Blue strain cells previously transformed with pRT66B and pATPR plasmids (described above). This culture was grown overnight at 37 ºC under constant stirring (around 180 rpm). Then, the culture was added to 300 ml of the same antibiotics-containing media and grown for another 4 hours under the same conditions. After this step, three one-liter flasks with fresh LB medium containing ampicillin (100 μg/ml) and kanamycin (50 μg/ml) were prepared. Then, 100 ml of the bacterial culture obtained in the previous step were added to each flask. These cultures were maintained under constant agitation at 37 ºC until the optical density at 600 nm reached a value between 0.6 and 0.8. Then, IPTG was added to a final concentration of 0.6 mM in order to induce the expression of the p66 subunit of the RT and the PR encoded within the plasmids. After maintaining the culture at 37 ºC for another 24 hours, the cells were harvested by centrifugation at 3000 rpm at 4 ºC for 30 min using a rotor JLA 10.500 (Beckman) and an Avanti J-26 XP centrifuge (Beckman Coulter).

3.2.1.3. Lysis of bacteria

The obtained bacterial pellet was resuspended in 20 ml of 40 mM Tris-HCl pH 8.0 buffer, containing 25 mM ethylenediaminetetraacetic acid (EDTA), 10% (w/v) sucrose, 2 mM phenylmethylsulfonyl fluoride (PMSF), 2 mM benzamidine, and 1 mg/ml lysozyme. The mixture was stirred at 4 ºC for 15 min. Then, another 20 ml of a solution containing 0.8% (w/v) NP-40, 20 mM dithiothreitol (DTT), 2 mM PMSF and 2 mM benzamidine were added and again mixed at 4 ºC for 15 min. Finally, we added NaCl to a final concentration of 0.5 M and the mixture was sonicated at a constant amplitude of 24 microns for 1 min using a Labsonic M sonifier (Sartorius). The sonicated sample was centrifuged at 15000 g at 4 ºC for 15 min using a rotor JA 25.50 and an Avanti J-26 XP centrifuge (Beckman Coulter). The resulting supernatant was diluted in 7 volumes of 50 mM sodium phosphate pH 6.8 buffer (made of 25 mM Na_2HPO_4 and 25 mM NaH_2PO_4) and stored at 4 ºC.

3.2.1.4. Cation-exchange chromatography

First of all, a cellulose phosphate P11 (Whatman) column was prepared. For that purpose, around 1.5 g of resin were washed in 25 volumes of 0.5 M NaOH during 5 min in agitation. After

allowing the resin to settle, the supernatant was removed and the washing process was repeated once again with 0.5 M NaOH, then twice with sterile water, twice with 0.5 M HCl and finally twice more with sterile water. Subsequently, the cellulose phosphate resin was packed into the column and was equilibrated for about 2 hours with 50 mM sodium phosphate pH 6.8 buffer, leaving the column at the same pH. The sample obtained in the previous section was passed through the column and then washed with around 200 ml of the equilibrating buffer.

The bound enzyme was eluted from the column using a cation concentration gradient prepared with 40 ml of 50 mM sodium phosphate pH 6.8 buffer, and 40 ml of the same buffer containing 2 M NaCl. The formation of the gradient was facilitated by a Hoefer™ SG100 gradient maker (Amersham Biosciences) containing a magnetic bar in the chamber used for the less concentrated dilution, and placed on a magnetic stirrer (ANS-001, BSB). Fractions of 30 drops per tube were collected in 2 ml tubes using a collector (FC205, Gilson). Afterwards, the presence of enzyme was checked by measuring the absorbance of all fractions at 280 nm using a Hitachi U-2000 spectrophotometer (Hitachi Ltd). The enzyme-enriched fractions were pooled and subjected to affinity chromatography.

3.2.1.5. Affinity chromatography

The pooled sample collected after the cation-exchange chromatography step was loaded in a 2 ml-column of Ni^{2+}-nitriloacetic acid-agarose (ProBond™, Invitrogen), previously equilibrated in 50 mM sodium phosphate pH 6.8 buffer containing 0.3 M NaCl. After loading the sample, the column was washed with 100 ml of 50 mM sodium phosphate pH 6.0 buffer containing 0.5 M NaCl. The RT interacts with the Ni^{2+} ions bound to the agarose matrix of the column through the imidazole groups of the C-terminal hexahistidine tail. The enzyme was eluted with an imidazole gradient, prepared with 20 ml of 50 mM sodium phosphate pH 6.0 buffer containing 0.5 M NaCl, and another 20 ml of the same buffer containing 0.5 M NaCl and 0.5 M imidazole. Fractions of 2 ml were collected and their absorbance at 280 nm was determined.

3.2.1.6. Dialysis and concentration

The RT-enriched fractions (around 20 ml) were pooled and subjected to dialysis using Visking® dialysis membranes of 16 mm and with a pore diameter of ~25 Å (Serva). Dialysis was performed in 250 ml of 50 mM Tris-HCl pH 7.0 buffer, containing 25 mM NaCl, 1 mM EDTA, 10% (w/v) glycerol, and 1 mM DTT added just before starting. Three changes of the dialysis solution were made, each one of at least 2 hours, maintaining one of them overnight. Finally, the dialyzed sample was concentrated to a final volume of 1-2 ml by ultracentrifugation at 3500 rpm and 4°C using

Centriprep® 30K (Merck Millipore Ltd) filter concentrators and a rotor JLA 16.250 in an Avanti J-26 XP centrifuge (Beckman Coulter). Ultracentrifugation usually involved at least two rounds of 30 min. Depending on the volume and RT concentration desired, the sample can also be concentrated by an Amicon® Ultra-4 Ultracel®-10K (Merck Millipore Ltd). Then, the sample was subjected to a microcentrifugation of 1 min in a Hettich Mikro200 to eliminate possible pellets. The concentrated RT aliquots obtained were stored at -20 ºC. The purity of the obtained heterodimers was determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein concentration was determined by using the RT molar extinction coefficient at 280 nm ($\varepsilon 280 = 260450\ M^{-1}\ cm^{-1}$), while the concentration of active enzyme was analyzed as a measure of the amount of RT bound to the template-primer at time zero, as described in Kati *et al*. (1992).

3.2.2. Purification of HIV-2 RTs

The purification of HIV-2 RTs relies on the co-expression of the large subunit (p68) of the RT and the HIV-2 protease (strain D194) in *E. coli* BL21(DE3)pLysS (**Table 3**), previously transformed with the plasmid pT5m (**Figure 10**), which was kindly provided by Amnon Hizi (Tel Aviv University, Israel). As an expression strain, BL21 has the potential advantage of being deficient in various proteases (e.g. *omp*T and *lon*) that could degrade recombinant proteins during purification. DE3 is a λ prophage derivative carrying the T7 RNA polymerase (gene 1) under the control of the IPTG-inducible *lac*UV5 promoter. The plasmid pT5m carries the coding sequence of the p68 subunit of HIV-2 RT as well as that of the HIV-2_{D194} PR, controlled by the T7 promoter. Finally, the plasmid pLysS allows to block basal expression of T7 RNAP by encoding the gene of the T7 lysozyme inhibitor. The sequence of this gene is found in the antisense orientation relative to the *tet* promoter, so only small amounts of T7 lysozyme are produced (**Figure 10**). Altogether, under IPTG induction and in the absence of glucose (that inhibits the formation of CAP/cAMP activators), T7 RNAP is expressed from the λDE3 prophage and drives the transcription of HIV-2 p68 RT-subunit and PR in the plasmid pT5m.

3.2.2.1. Site-directed mutagenesis

The RT-coding region of the p68 subunit of HIV-2_{ROD}, and the PR-coding region of HIV-2_{D194} are found in the polycistronic plasmid pT5m under the control of a single T7 promoter (**Figure 10**). The resulting long polycistronic mRNA contains both genes, each one having its own ribosome-binding-site (RBS) and stop codon. Six amino acids of histidine flank both, N- and C-termini of p68. It has been described that HIV-2 RTs with His_6 tags only at the carboxyl terminus of p68 (like in HIV-1 RT) did not bind to NTA-agarose-columns, and then they could not be purified (Sevilya *et*

al., 2001). In our case, the N-terminal His_6 tag is linked to the RT through an amino acid sequence that contains a thrombin cleavage site. The vector pT5m was used as template to introduce the amino acid substitution K65R and the combination of mutations K65R/Q151M/M184V using the "QuickChange™ Site-Directed Mutagenesis Kit" of Stratagene according to the manufacturer's instructions. Mutagenic primers are shown in **Table 6**.

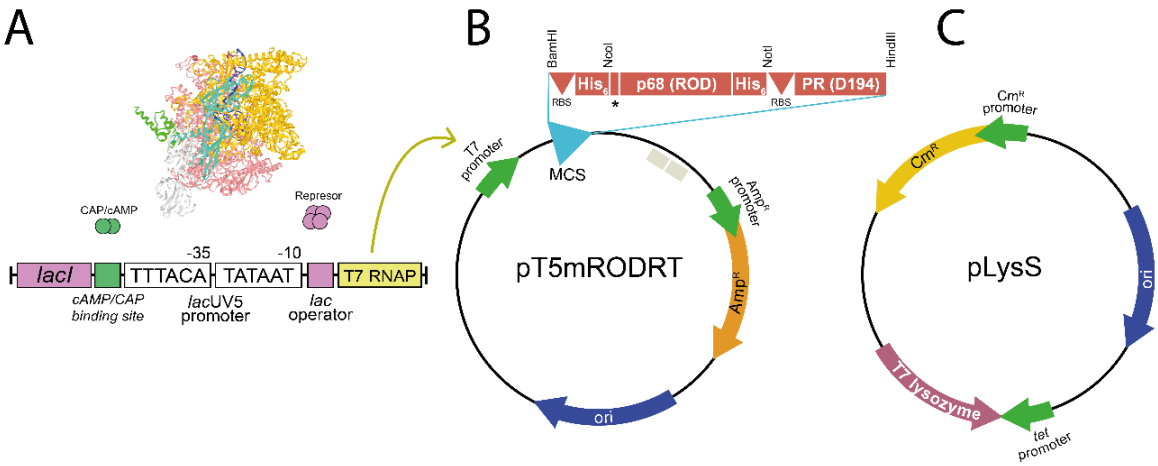


Figure 10. Plasmids and genes involved in HIV-2 RTs purification. (A) λDE3 prophage derivative. The prophage is found within BL21 bacteria and encodes the T7 RNAP (gene 1) under the control of an IPTG-inducible promoter. Under activating conditions (absence of glucose, promoting the formation of CAP/cAMP stimulators) and in the presence of IPTG, the *E. coli* RNAP drives the transcription of T7 RNAP gene, thereby allowing its expression. **(B)** Plasmid pT5m. This plasmid leads the synthesis of a polycistronic messenger RNA, which has two ribosome-binding sites (RBS), one upstream of the p68 RT-subunit gene and the other upstream of the HIV-2_{D194} PR gene. Sequences coding six-histidine tags are present at the amino and carboxyl termini of p68 subunit (Sevilya *et al*., 2001). The asterisk represents a sequence that once translated acts as a recognition site for digestion with thrombin. BamHI and HindIII restriction sites delimit the whole insert within the pT5m plasmid. A T7 promoter drives the transcription of the whole construction. **(C)** Plasmid pLysS. This plasmid encodes the T7 lysozyme in an antisense orientation, blocking the basal expression of T7 RNAP mediated by the λDE3 prophage. Amp^R and Cm^R indicate the regions codifying for the ampicillin and chloramphenicol resistance genes, respectively. Ori, indicates the origin of replication.

Table 6. Primers used in HIV-2_{ROD} RT mutagenesis reactions

Mutation	Mutagenic primer sequences
K65R	3'-CAAAAACAGGAAAGAGAACTAACGTTTA-5' 5'-GTTTTTGTCCTTTCTCTTGATTGCAAAT-3'
Q151M	3'-AGGGAAGGTAGGGTAACCGTTCTGAAAT-5' 5'-TCCCTTCCATCCCATTGGCAAGACTTTA-3'
M184V	3'-TCTATAGTAGGTGCATGACTTACTATTA-5' 5'-AGATATCATCCACGTACTGAATGATAAT-3'

Left column indicates the mutation(s) introduced in the RT. Mutagenized codons are underlined.

Mutations found in the triple mutant K65R/Q151M/M184V were orderly added starting from mutant K65R, then introducing mutation Q151M, and finally M184V. The mutagenesis reaction

was performed as described in section 3.2.1.1, and the product was transformed in *E. coli* DH5α competent cells by thermal shock. As described above, ampicillin-resistant colonies were selected and the desired RT sequence confirmed by sequencing. After ratifying the presence of the mutation, the plasmid pT5m was used to transform the *E. coli* BL21(DE3)pLysS strain by thermal shock. Finally, ampicillin- and chloramphenicol-resistant colonies were selected for the following step.

3.2.2.2. Expression of HIV-2 RT p68/p54 heterodimers

Four flasks with 40 ml of LB medium containing 40 mM glucose, 100 μg/ml ampicillin and 34 μg/ml chloramphenicol were inoculated with 5-10 μl of a glycerol-containing stock of *E. coli* BL21(DE3)pLysS cells previously transformed with the pT5mRODRT plasmid (described above). These cultures were grown overnight at 30 ºC under constant stirring (ca. 180 rpm). After this step, each 40 ml-flask of bacterial culture was added to a one-liter flask with fresh antibiotics-containing medium, supplemented with 2 mM glucose. The four cultures were maintained under constant agitation at 30 ºC, until the optical density at 600 nm reached a value of 0.7. Therefore, IPTG was added to a final concentration of 1 mM, in order to induce the expression of the p68 subunit of the RT and the PR encoded within the plasmid pT5m. After maintaining the cultures at 30 ºC for another 3 hours, cells were harvested by centrifugation at 3500 rpm at 4 ºC for 30 min using a rotor JLA 10.500 and an Avanti J-26 XP centrifuge (Beckman Coulter).

3.2.2.3. Lysis of bacteria

The pellets obtained in the four cultures were re-suspended together in 24 ml of 50 mM sodium phosphate pH 8.0 cold buffer, containing 50 mM NaCl, 0.75 mg/ml hen egg-white lysozyme and 1.5 mM PMSF. After keeping the mixture on ice stirring for 25 min, 1.72 ml of 4 M NaCl were added. Then, the solution was subjected to brief periods of sonication at a constant amplitude of 24 microns using a Labsonic M sonifier (Sartorius). Afterward, the sample was centrifuged at 23000 rpm for 60 min at 4°C using a Beckman JA 25.50 rotor and an Avanti J-26 XP centrifuge (Beckman Coulter). The obtained supernatants were diluted with one volume of 66 mM sodium phosphate pH 7.0 buffer, containing 0.3 M NaCl, and were stored at 4°C.

3.2.2.4. Affinity chromatography

A 3-ml Ni^{2+}-nitriloacetic acid-agarose column (Invitrogen, Carlsbad, CA, USA), was prepared and equilibrated with 50 mM sodium phosphate pH 7.0 buffer containing 0.3 M NaCl. After applying the supernatant sample, the column was washed with 50 ml of the equilibrating buffer and with 90 ml of 50 mM sodium phosphate pH 6.0 buffer containing 0.3 M NaCl, 10% glycerol and 20

mM imidazole, in order to eliminate unspecific binding. Subsequently, the RT was eluted with a 20 mM–0.5 M imidazole gradient, prepared with 30 ml of 50 mM sodium phosphate pH 6.0 buffer containing 0.3 M NaCl, and 10% glycerol. Absorbance at 280 nm was measured for every fraction and those showing the highest values were analyzed by SDS-PAGE. Fractions containing the heterodimer were pooled and dialyzed against 250 ml of 50 mM sodium phosphate pH 6.8 buffer, using Visking® dialysis membranes of 16 mm and with a pore diameter of ~25 Å (Serva). Three changes of the dialysis solution were made, maintaining each one for at least 2 hours.

3.2.2.5. Cation-exchange chromatography

The dialyzed sample was applied on a 5 ml P11 phosphocellulose column (Whatman, Maidstone, Kent, England), previously equilibrated with the dialysis buffer. After washing the column with 50 mM sodium phosphate pH 6.8 buffer containing 0.4 M NaCl, the RT was eluted with a total gradient of 60 ml from 0.4 to 1.2 M NaCl prepared in the washing buffer (i.e., 50 mM sodium phosphate pH 6.8 buffer containing 0.4 M NaCl). Fractions eluted from the salt gradient were analyzed by spectrophotometry at 280 nm and SDS-PAGE, and those containing the heterodimer were pooled and dialyzed against 50 mM Tris-HCl pH 7.0, containing 25 mM NaCl, 1 mM EDTA pH 8, 1 mM DTT and 10% glycerol. Again, three changes of the dialysis solution were made, each one of at least 2 hours, and maintaining one of them overnight. Finally, the enzyme was concentrated in Centriprep® 30K and Amicon® Ultra-4 Ultracel®-10K (Merck Millipore Ltd) concentrators to <0.5 ml, and stored at −20 °C. Enzymes were quantified by active site titration before biochemical studies, as described in Kati *et al*. (1992).

3.3. Purification of T7 RNA polymerase

The bacteriophage T7 RNAP was purified as an N-terminal 6-histidine tagged monomer after its expression in *E. coli* BL21(pREP4) strain (**Table 3**), previously transformed with plasmid pQE9T7 (**Figure 11**). The T7 RNAP-coding sequence was cloned in a pQE9 plasmid in sites generated by BamHI and HindIII endonucleases, leading to an in-frame translational fusion of the polymerase with the 6-histidine tag codified in the vector (**Figure 11**) (Ellinger and Ehricht, 1998). The resulting plasmid pQE9T7 was kindly provided by Ralf Ehricht (Alere Technologies Gmbh, Jena, Germany). The plasmid pREP4 found within BL21 bacteria constitutively expresses the *lac* repressor protein (encoded by the *lac*I gene) that trans-represses the bacteriophage T5 promoter found on the pQE9T7 expression vector. Under the addition of IPTG, the promoter T5 is induced and transcription catalyzed by the *E. coli* RNAP initiates.

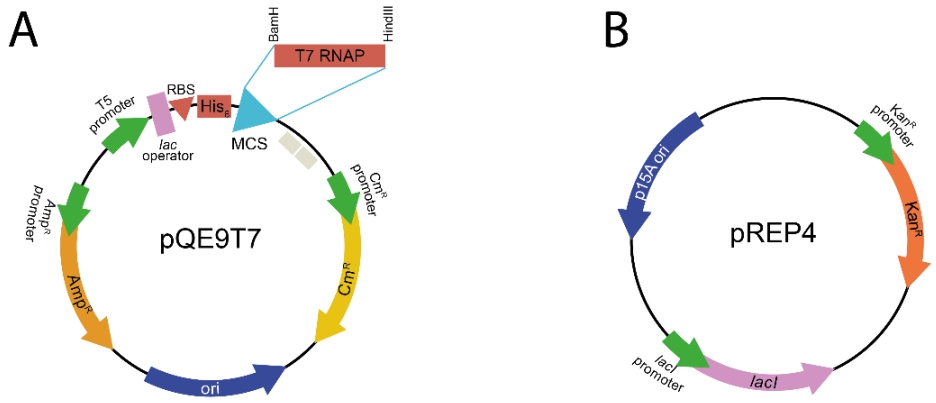


Figure 11. Plasmids used in T7 RNAP purification. (A) Plasmid pQE9T7. This plasmid contains the T7 RNAP-coding sequence cloned in the restriction sites of BamHI and HindIII endonucleases. The polymerase sequence is arranged in such a way that leads to an in-frame translational fusion of the T7 RNAP with the 6-histidine tag codified in the vector (Ellinger and Ehricht, 1998). **(B)** Plasmid pREP4. It expresses constitutively the *lac* repressor protein (encoded by the *lac*I gene) that trans-represses the bacteriophage T5 promoter found on the pQE9T7 vector. Under the addition of IPTG, the promoter T5 is induced and transcription initiates via *E. coli* RNAP, allowing the expression of T7 RNAP. Amp^R, Cm^R and Kan^R indicate regions codifying for the ampicillin, chloramphenicol and kanamycin resistance genes, respectively. Ori, indicates the origin of replication.

3.3.1. Expression of T7 RNAP

A fresh colony of *E. coli* BL21(pREP4) strain cells previously transformed with the plasmid pQE9T7 was inoculated in 20 ml of Tryptone Yeast medium (TY medium contains 6 g bactotriptone, 3 g yeast extract and 0.38 g $CaCl_2$ per liter). Antibiotics ampicillin and kanamycin were added at 100 µg/ml and 25 µg/ml, respectively. This culture (20 ml) was incubated at 37 °C overnight under constant agitation. Then, the culture was added to a 1-liter flask of fresh medium containing ampicillin/kanamycin, and maintained again at 37 °C under constant agitation until an optical density at 600 nm of 1.0 was reached. Subsequently, the expression of T7 RNAP was induced with 1 mM IPTG during three additional hours and the cells were harvested by centrifugation at 3600 g during 30 min at 4°C using a rotor JLA 10.500 and an Avanti J-26 XP centrifuge (Beckman Coulter).

3.3.2. Lysis of bacteria

The bacterial pellet was resuspended in 15 ml of 50 mM sodium phosphate pH 8.0, containing 300 mM NaCl and 0.5 mM DTT. After adding lysozyme, PMSF and benzamidine to a final concentration of 6 mg/ml, 0.5 mg/ml and 2.5 mM, respectively, the mixture was stirred on ice for 30 min and sonicated at a constant amplitude of 24 microns for 40 s using a Labsonic M sonifier (Sartorius). Then, the sample was centrifuged at 15000 g at 4 ºC for 30 min using a rotor JA 25.50 and an Avanti J-26 XP centrifuge (Beckman Coulter). The resulting supernatant was collected for the following step and stored at 4°C.

3.3.3. Affinity chromatography

The supernatant containing His_6-tagged T7 RNAP was applied twice on a 6-ml Ni^{2+}-NTA Sepharose column previously equilibrated with 50 mM sodium phosphate buffer pH 8.0 containing 300 mM NaCl and 0.5 mM DTT. Then, the column was washed with 100-150 ml of the same buffer, containing 10% glycerol, 20 mM imidazole and 20 µg/ml PMSF. The protein was eluted by a 60-ml linear 0.02-0.5 M imidazole gradient in 50 mM sodium phosphate pH 8.0, containing 300 mM NaCl and 0.5 mM DTT. Fractions were measured spectrophotometrically at 280 nm and those showing a peak were examined by SDS-PAGE.

3.3.4. Dialysis and concentration

Fractions containing T7 RNAP were pooled and dialyzed against 250 ml of 20 mM sodium phosphate buffer pH 7.7, containing 100 mM NaCl, 1 mM EDTA, 10% glycerol, and 1 mM DTT added just before dialysis, using Visking® dialysis membranes of 16 mm and with a pore diameter of ~25 Å (Serva). Four dialysis of 2 hours each were carried out, including one maintained overnight. The volume (7 ml) was then reduced by centrifugation at 1100 g and 4°C twice during 15 min and once more during 5 min, using Centriprep® 30K filter concentrators. The protein concentration in the final preparation was determined by using the T7 RNAP molar extinction coefficient at 280 nm ($\varepsilon 280 = 1.4 \times 10^5\ M^{-1}\ cm^{-1}$) (King *et al*., 1986), as well as according to the method of Bradford (Bradford, 1976). The concentrated T7 RNAP aliquots obtained were stored at −20 ºC.

3.4. M13mp2 forward mutation assays

3.4.1. Determination of DNA-dependent DNA synthesis fidelity of RTs

3.4.1.1. Preparation of “gapped DNA”

Twelve ml of 2xYT medium (containing 16 g bactotriptone, 10 g yeast extract and 10 g NaCl, per liter) were inoculated with a fresh colony of *E. coli* NR9099 strain (**Table 3**), previously grown on M9 minimal plates. The culture was maintained at 37 °C under constant stirring for about 4 hours until the optical density at 600 nm reached a value of 0.6. Then, this initial culture was added to a 1-liter of fresh medium and maintained for another two hours. After inoculating an isolated plaque of the WT M13mp2 phage, the culture was grown overnight at 37 °C under constant agitation.

Afterwards, the sample was centrifuged at 4000 rpm during 30 min at 4 °C using a rotor JLA-10.500 and an Avanti J-26 XP centrifuge (Beckman Coulter). Subsequently, the supernatant was separated from the pellet. The supernatant contains the cell-free M13mp2 filamentous phage, whose

genome consists of a circular positive-polarity ssDNA, while the pellet contains the replicative form (RF) of the phage as dsDNA within the bacteria. Thus, the supernatant was used for the purification of the M13mp2 ssDNA, while the dsDNA was extracted from the bacterial pellet (described below). An illustrative representation of the whole assay is shown in **Figure 12**.

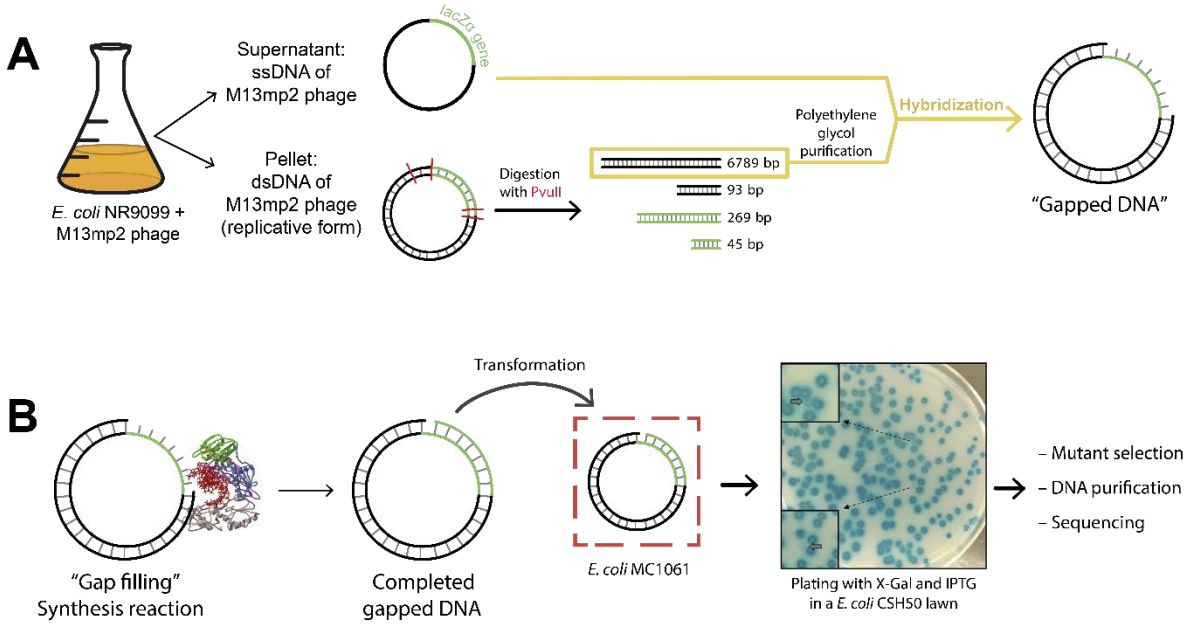


Figure 12. Measurement of DNA-dependent DNA synthesis fidelity based on M13mp2 forward mutation assays. (A) Gapped DNA formation. Gapped DNA was obtained by hybridizing the M13mp2 ssDNA with a 6789-bp dsDNA fragment. This fragment lacking the *lacZα* sequence was obtained by polyethylene glycol purification after digestion of the M13mp2 dsDNA (or replicative form) with PvuII. **(B)** Gapped DNA serves as substrate for DNA-dependent DNA synthesis reactions mediated by RTs. The product of the "gapped filling" reaction was used to transform *Escherichia coli* MC1061 competent cells. The transformed cells were plated onto M9 medium-containing plates with X-Gal and IPTG with *E. coli* CSH50 cells. Errors made by RTs while copying DNA templates result in a decrease in α-complementation and could be detected by the altered color phenotype (light blue or colorless) of the mutant plaques. After mutant selection and DNA purification, the error specificity was determined by sequencing.

3.4.1.1.1. Purification of phage M13mp2 ssDNA

The supernatant isolated in the previous step was centrifuged once more under the same conditions in order to ensure that bacteria were completely eliminated. Then, supernatant was divided into four 250 ml aliquots and each one was subjected to the following steps. Fifty ml of 15% polyethylene glycol (PEG) 8000 and 2.5 M NaCl were added. The mixture (300 ml) was inverted several times, maintained on ice during 10 min, and centrifuged at 4000 rpm for 30 min at 4 °C using a rotor JLA 10.500. The pellet containing the precipitated phage was resuspended in a mixture made of 2 ml of 1 M Tris-HCl buffer pH 8.0, 2 ml of 0.5 M EDTA and 14 ml of distilled water, and kept on ice for 30 min. Again, the sample was centrifuged at 2500 rpm for 15 min at 4 °C using a rotor JA-25.50 to eliminate residual PEG. This rotor was used in subsequent ultracentrifugations.

Afterwards, 100 µl of proteinase K (10 mg/ml) (New England Biolabs) were added to the supernatant in order to degrade the ~2700 copies of major coat protein (pVIII) that cover each genome particle (Stopar *et al*., 2003; Houbiers and Hemminga, 2004; Kehoe and Kay, 2005). The mixture was stirred and incubated at 37 °C for 30 min under constant agitation. Solutions were pooled two by two, and four ml of a cationic detergent [5% -(w/v)- cetyltrimethylammonium bromide and 0.5 M NaCl] were added to each tube containing two mixtures. After incubating 10 min at room temperature, the samples were centrifuged at 12000 g for 30 min at 4 °C. The pellets were resuspended in 6 ml of 1.2 M NaCl, and 5 ml of 100% isopropanol were added, maintaining the mixture overnight at −20°C. Then, the samples were centrifuged at 12000 g for 30 min at 4 °C and the ssDNA-containing pellets were precipitated with ethanol. After resuspension in 500 µl of distilled water, the ssDNA was checked in 1% agarose gels and kept at −20°C.

3.4.1.1.2. Purification of phage M13mp2 dsDNA and digestion

The M13mp2 dsDNA was purified from the pellet (obtained as described in section 3.4.1.1), using the commercial kit "Pure Yield™ Plasmid Maxiprep System" (Promega) following the manufacturer's instructions. Then, 100 µg of dsDNA were digested with the restriction enzyme PvuII (New England Biolabs) in 100-200 µl of 10 mM Tris-HCl buffer pH 7.9, containing 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT, at 37 °C during 2 hours. One unit of PvuII per 1.5 µg of dsDNA was used. The RF contains three restriction sites for PvuII, providing four fragments of dsDNA of 6789, 268, 93 and 46 base pairs (bp) (**Figure 12**). After precipitation with ethanol, the digestion was analyzed in a 1% agarose gel.

The digested DNA sample was then adjusted to 6% PEG 8000 and 0.55 M NaCl, leaving the DNA at a concentration ranging from 0.5 to 0.05 µg/µl. The mixture was incubated overnight at 4 °C and later centrifuged at 14000 rpm for 20 min at 4 °C in a microcentrifuge (Mikro200, Hettich). 6% PEG promotes the precipitation of fragments of more than 1650 bases (Lis, 1980), thereby allowing to pellet in the centrifugation exclusively the fragment of 6789 bp. The pellet was then resuspended in 20 µl of 0.2 M NaCl and later precipitated by ethanol, in order to remove residual PEG. After resuspension, the 6789-bp dsDNA fragment was checked in a 0.8% agarose gel.

3.4.1.1.3. Hybridization of ssDNA and the 6789-bp dsDNA fragment

Five µg of the 6789-bp dsDNA fragment were diluted in 50 µl (0.1 µg/µl) in order to reduce the ionic strength and facilitate the separation of the strands. After an incubation of 5 min at 70 °C, the ssDNA was added in a 1:1 proportion (taking into account the ssDNA co-purified with the dsDNA sample) and was maintained at 70 °C for 2 min. Then, the mixture was kept on ice for 5 min

and saline-sodium citrate (SSC) 20X buffer (3 M NaCl and 300 mM sodium citrate adjusted to pH 7.0) was added until a final concentration of 2X. Then, the sample was heated 5 min at 60 °C and kept on ice 10 min. The formation of the gapped DNA was analyzed in a 0.8% agarose gel (of dimensions 15 cm x 25 cm, or 250 ml) at 90 V during 3.5 h. Fractions containing properly annealed gapped DNA were pooled and precipitated with ethanol. After resuspension in water, the concentration of the product was measured in a NanoDrop ND-1000 (Thermo Scientific) apparatus and the proportion between gapped DNA and dsDNA was estimated in agarose (0.8%) gel electrophoresis to determined the genuine concentration of gapped DNA.

3.4.1.2. Gap-filling synthesis reaction

"Gap-filling" synthesis reactions were carried out at 37 ºC during 60 min, in 25 µl of 25 mM Tris-HCl buffer pH 8.0, containing 100 mM KCl, 2 mM DTT, 4 mM $MgCl_2$, 250 µM of each dNTP and 100 nM RT. Each reaction contained 100-150 ng of gapped DNA, which was preincubated at 55 ºC during 5 min to reduce secondary structure in the ssDNA and avoid undesirable hybridizations. Nine or ten gap filling reactions were performed for each tested RT (that from WT HIV-1_{ROD}, and mutant RTs $K65R_{ROD}$ and $K65R/Q151M/M184V_{ROD}$). The reactions were stopped by the addition of EDTA at a final concentration of 6 mM. In order to check if the synthesis had been completed, the obtained products were analysed on a 0.8% agarose gel (of dimensions 15 cm x 25 cm, or 250 ml) at 65 V during 16-18 h. At least 22 µl of sample should be applied to assure a proper visualization. Gapped DNA was used as a control, in order to compare its electrophoretic mobility with that of the "gap-filling" product (difference of 407 nucleotides). Samples clearly filled were pooled together.

3.4.1.3. Preparation of competent cells

A fresh colony of *E. coli* MC1061 cells was inoculated in 10 ml of 2xYT medium and was grown at 37°C overnight under constant agitation. This initial culture was added to 500 ml of 2xYT, and one hour later, 2xYT medium was added to complete a total volume of 1 L. The culture was grown under the same conditions until the absorbance at 600 nm achieved a value between 0.5 and 0.8. The volume was then kept on ice during 30 min and centrifuged at 4000 g and 4 °C for 30 min using a rotor JLA-10.500. The pellet was washed in one liter of sterile cold water and centrifuged at 2200 g and 4 °C for 20 min. The process was repeated with 500 ml of sterile cold water. Then, the pellet was resuspended in 20 ml of 10% cold glycerol and the last centrifugation was carried out at 3000 g and 4 °C for 15 min using a rotor JA 25.50. After resuspending the bacteria in 3 ml of 10% glycerol, 50 µl aliquots were prepared, and immediately frozen by using dry ice and ethanol. Competent bacteria were stored at −80 °C.

3.4.1.4. Electroporation and plating

The product of the gap filling reaction (1 µl), obtained as described in section 3.4.1.2, was inoculated in 50 µl of MC1061 competent cells. This bacterial strain is preferred because it shows a very high electroporation efficiency (Bebenek and Kunkel, 1995). After keeping the sample on ice for 3 min, the mixture was transferred to an electroporation cuvette (0.2 cm, BioRad) and electroporated at 200 Ω resistance, 25 µF capacitance and 2.5 kV voltage using a "Gene Pulser" (BioRad). Immediately afterwards, 500 µl of SOC medium (20 g/l bactotriptone, 5 g/l yeast extract, 0.5 g/l NaCl, 2.5 mM KCl, 20 mM $MgCl_2$, 20 mM $MgSO_4$, and 20 mM glucose) were added. Comparing the resulting time constant with that of bacteria without DNA (around 5 ms), the amount of sample to use in the plating can be estimated.

The mixture of plating contains 2.5 ml of "soft-agar" (0.9% NaCl and 0.8% agar) preheated at 49 °C, 50 µl of 5% (w/v) cold X-Gal, 25 µl of 0.1 M IPTG, 500 µl of a logarithmic culture of bacteria *E. coli* CSH50, and 25-75 µl of the electroporated sample (the exact amount is determined in initial trials). The whole mix is stirred and poured on M9 minimal plates. After solidifying, plates are incubated upside down overnight at 37 °C. During that time, M13mp2 phages are produced in MC1061 bacteria and infect the CSH50 lawn cells, which can grow in the minimum medium plates. Alpha-complementation takes place in CSH50 bacteria, between the α fragment of the β-galactosidase enzyme carried by phages and their Ω fragment.

3.4.1.5. Selection and analysis of mutant plaques

The infection produces well-defined lysis plaques, including WT and different types of mutants. WT plaques were arbitrarily assigned a color value of 4+ (dark blue), while mutant plaques vary from almost WT (3+) to colorless (0+), going through medium blue (2+) and light blue (1+) (**Figure 13**) (Bebenek and Kunkel, 1995). This color gradation in plaque phenotype is a consequence of the errors introduced by the RT while synthesizing the DNA during the "gap-filling" reaction. Thus, these errors provoke a partial or complete loss of β-galactosidase activity after the α-complementation. It is essential to carefully examine the plates 24 h following plating, since the 3+ phenotype is more easily distinguished at this time; but also several days later when plaques have acquired more intensity. Mutants were collected and stored at 4 °C in 1 ml of 0.9% NaCl. After selecting mutants, every single plaque is counted because mutant frequencies were calculated as the ratio of mutant plaques to the total number of plaques screened. In case of doubt, the possible mutant plaque is plated together with a WT by using only 1 µl of a dilution 1/500 of this mix. Then, if all lysis plaques obtained in this verification plating have the same color (**Figure 13E**), the plaque was in fact a WT, whereas if two different colors can be clearly observed (**Figure 13A-D**), the lysis plaque under study was confirmed as a mutant.

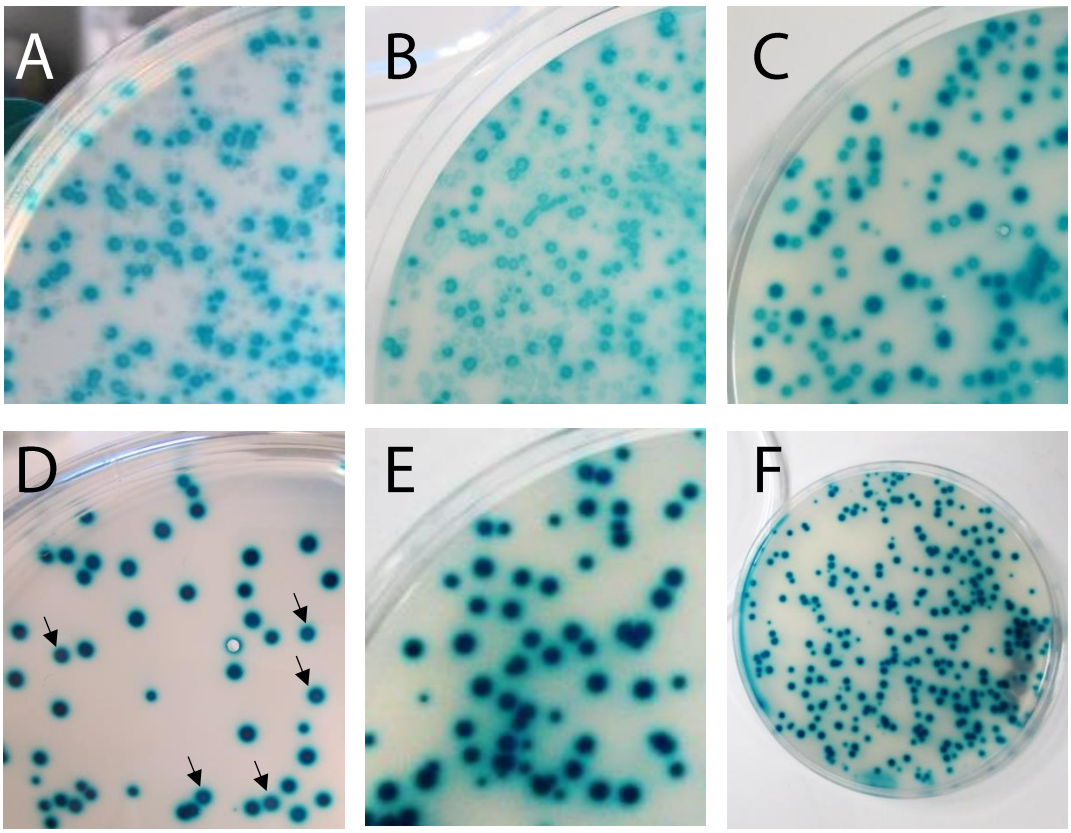


Figure 13. WT and mutant lysis plaques obtained in M13mp2-based forward mutation assays. (A-D) Mutant plaques together with a WT, whose phenotypes vary within a shade of blue tonalities, with an arbitrarily assigned color value of 0+ (A, colorless), 1+ (B, light blue), 2+ (C, blue), 3+ (D, dark blue, almost WT). **(E)** WT lysis plaque. **(F)** Prototypic plating obtained in an assay.

In order to purify the phage dsDNA, 100 µl of phage-containing NaCl in 2xYT medium were grown in agitation overnight at 37 °C and DNA was extracted from bacterial pellet by using the commercial kit "Wizard® Plus SV Minipreps DNA purification System" (Promega). As a verification procedure, 200 ng of mutant DNA were digested with 1.5 units of MscI and 10 units of BamHI (New England Biolabs), and only those mutants fitting the correct pattern of bands were sent for sequence analysis. M13mp2 DNA has one restriction site per each enzyme, resulting in two fragments of 2860 and 4336 bp. The digestion was carried out in 10 µl of 20 mM Tris-acetate buffer pH 7.9, containing 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT and 0.1 mg/ml BSA. The product was analyzed in a 1% agarose gel. Then, mutant phenotypes were confirmed by nucleotide sequencing of the phage dsDNA using primer 5′-GCTTGCTGCAACTCTCTCAG-3′ (Macrogen Inc.).

Specific error rates were calculated as described by Bebenek and Kunkel (1995). Briefly, the number of a particular class of errors is divided by the total number of plaques observed in the assay, by the number of sites at which the error under consideration can be detected and by 0.6 (the probability that a polymerase error in the newly synthesized minus strand will be expressed in *E.*

coli). Errors were detected at 125 template positions for base substitutions and 148 for frameshifts (79 at runs, and 69 at non-runs). Overall error rates were determined by considering 237 template positions were any phenotypic change could be detected (**Figure 14**). It should be noted that silent mutations and mutations that do not affect the β-galactosidase activity may not be detected in these assays. It has been estimated that this limitation results in a 2- to 3-fold underestimation of the actual mutation rates (Abram *et al*., 2010).

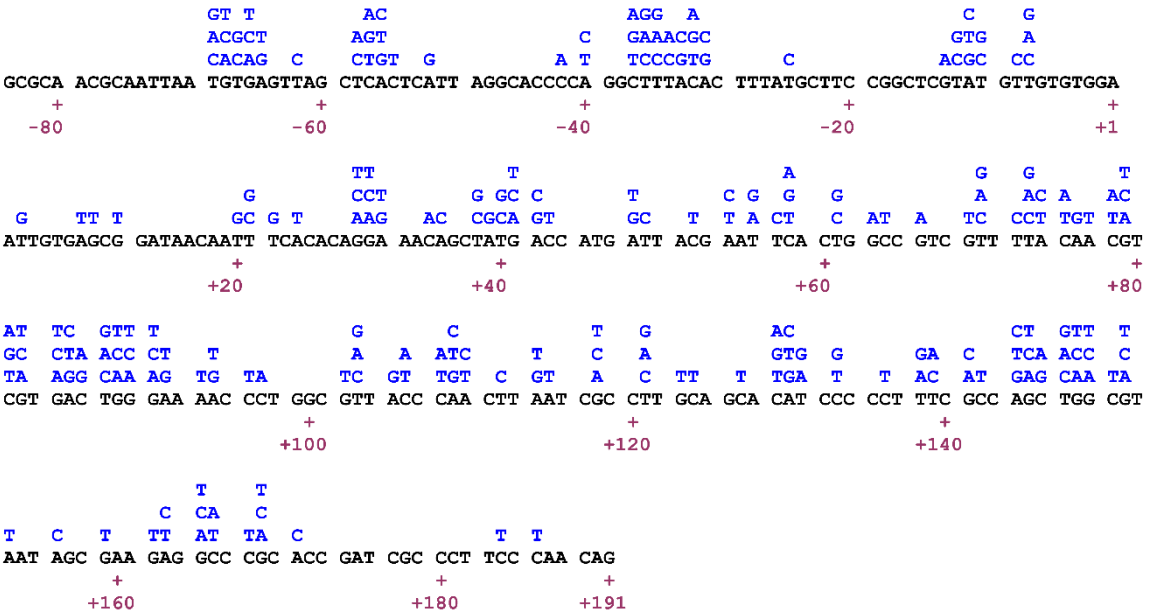


Figure 14. Base substitutions causing a phenotypically detectable change in *lacZα* target. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the wild-type template sequence of the *lacZα* target (Bebenek and Kunkel, 1995).

3.4.2. Determination of RNA-dependent DNA synthesis fidelity of RTs

3.4.2.1. Purification of RNA template

A modified M13mp2 phage was used for the study of RT's fidelity on RNA templates (the assay is depicted in **Figure 15**). In that phage, the promoter of the bacteriophage T7 RNAP was placed upstream the *lacZα* gene, between bases −112 and −113 (Boyer *et al*., 1992). The replicative form containing the T7 RNAP promoter (RF/T7) was obtained as described in sections 3.4.1.1 and 3.4.1.1.2 for the RF. Around 40 µg of RF/T7 were digested with 1 unit of FspI per µg of DNA, in 50 µl of 20 mM Tris-acetate buffer pH 7.9, containing 50 mM potassium acetate, 10 mM magnesium acetate and 1 mM DTT. The reaction was incubated at 37 °C during 2 h, followed by a standard extraction with phenol:chloroform:isoamyl alcohol (25:24:1, by volume) to eliminate the enzyme, and a precipitation with ethanol to remove salts. The linearized DNA was resuspended in 20 µl of RNase-free water. The simple restriction site for FspI is found at position +195 of the *lacZα* gene.

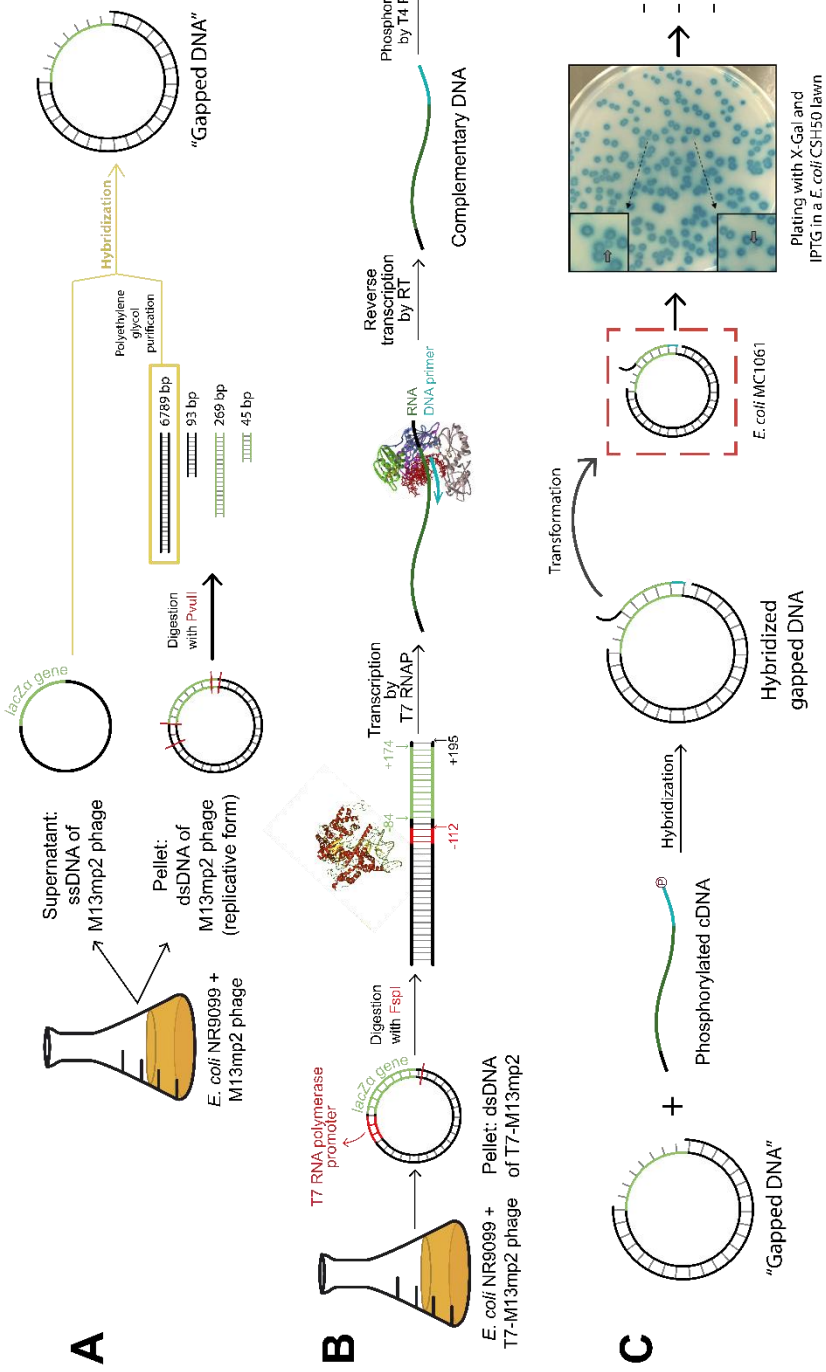


Figure 15. Measurement of RNA-dependent DNA synthesis fidelity based on M13mp2 forward mutation assays. (A) Gapped DNA formation. Gapped DNA was obtained by hybridizing the M13mp2 ssDNA with a 6789-bp dsDNA fragment. This fragment lacking the *lacZα* sequence was obtained by polyethylene glycol purification after digestion of the M13mp2 dsDNA (or replicative form) with PvuII. **(B)** Synthesis of phosphorylated cDNA. Simultaneously, the dsDNA of a modified M13mp2 phage containing the promoter for T7 RNAP upstream the *lacZα* gene (T7-M13mp2) was purified. After its digestion with FspI, transcription with T7 RNAP produces a 313-nt *lacZα* transcript. This RNA was used as template for reverse transcription by different RTs, providing a cDNA product that was then phosphorylated. **(C)** Gapped DNA and phosphorylated cDNA were hybridized and the product was used to transform *Escherichia coli* MC1061 competent cells. The transformed cells were plated onto M9 medium-containing plates with X-Gal and IPTG with *E. coli* CSH50 cells. Errors made by RTs during reverse transcription result in a decrease in α-complementation and could be detected by the altered color phenotype (light blue or colorless) of the mutant plaques. After mutant selection and DNA purification, the error specificity was determined by nucleotide sequencing.

Five or six µg of linearized RF/T7 were used as template for transcription by T7 RNAP. Reactions were carried out in a total volume of 50 µl. Two different transcription conditions were used: (i) 40 mM Tris-HCl buffer (pH 7.9), containing 6 mM $MgCl_2$ and 0.5 mM of each NTP (optimal conditions); and (ii) 40 mM Bis-Tris or PIPES buffer (pH 6.75), containing 1.5 mM $MgCl_2$ and 0.2 mM of each NTP. In both cases, reactions also contained 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 50-60 units of RNasin® Plus RNase inhibitor (Promega). For these experiments, 40 units of T7 RNAP (Promega) or 150 nM of recombinant T7 RNAP were used. Samples were incubated at 37 °C for 2 hours to obtain a 313-nucleotide RNA product (corresponding to positions -118/+195 in the *lac*Z gene). The dsDNA RF/T7 template was then digested with 5-6 units of RQ1 RNase-free DNase (Promega) at 37 °C during 15 min, and the RNA product was purified by extraction with phenol:chloroform:isoamyl alcohol (25:24:1, by volume), followed by ethanol precipitation in the presence of sodium acetate. The size and integrity of the RNA was tested in a 1% agarose gel, in comparison with RNA markers of 0.1-1.0 kb (Perfect RNA Novagen).

3.4.2.2. Synthesis of complementary DNA

The purified RNA was used as template in reverse transcription reactions carried out with WT AMV, MLV, HIV-1_{BH10} and HIV-1_{ESP49} RTs, and mutant RTs $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$. The cDNA synthesis reactions were carried out at 37 °C for 2 hours in 50 µl of 25 mM Tris-HCl buffer pH 8.0, containing 100 mM KCl, 2 mM DTT, 4 mM $MgCl_2$, 40 units of RNasin® Plus RNase inhibitor (Promega), 250 µM of each dNTP and 100 nM RT (active site concentration). Each reaction contained 2 pmol of RNA, previously heated to 65-70 °C for 5 min, and primed with a two-fold molar excess of the Rtr174/18 oligonucleotide (5'-CTGTTGGGAAGGGCGATC-3'). Six to ten reverse transcription reactions were performed for each enzyme. The reactions were stopped by adding EDTA at a final concentration of 15 mM, and incubated at 80 °C for 5 min to denature the hybrid cDNA-RNA.

The RNA template used in reverse transcription was digested at 37 °C for 1 hour, after adding 1 µl of RNase Cocktail™ containing RNase A and RNase T1 (Ambion). Two tubes of the obtained cDNA products were pooled together. Subsequently, the sample was purified by phenol extraction, precipitated with ethanol and resuspended in 10 µl. Simultaneously, another cDNA synthesis reaction was performed with 1 µCi of α[^{32}P]dCTP (Perkin Elmer) only as a control of the concentration and size of the product (309 bases). The labeled cDNA was applied on a 12% polyacrylamide gel with 8 M urea and subjected to electrophoresis at 70W 3000V 50-100 mA for about 4 hours. Then, the gel was revealed through exposition to phosphoroimaging plates (Fujifilm BAS-MP 2040S) and scanned with a BAS 1500 (Fujifilm).

3.4.2.3. Phosphorylation and hybridization of cDNA with "gapped DNA"

Ten µl of cDNA were phosphorylated with 5 units of the bacteriophage T4 polynucleotide kinase (PNK) (New England Biolabs) and 1 mM of ATP in 12.7 µl of 70 mM Tris-HCl buffer pH 7.6, containing 10 mM $MgCl_2$ and 5 mM DTT. After 1-h incubation at 37 °C, the reaction was heat-inactivated at 65 °C for 5 min. The cDNA was then hybridized with 8 µl of 2 nM gapped DNA, previously heated during 5 min at 55 °C. Two µl of SSC 20X were added to facilitate the process. The cDNA was in excess over gapped DNA and the mixture was heated to 70-75 °C for 5 min and slowly cooled at room temperature. Three µl of each hybridized product where separated and the rest was analyzed by electrophoresis in 0.8% agarose gels, as indicated in section 3.4.1.2.

3.4.2.4. Plating, selection and analysis of mutant plaques

Preparation of competent bacteria, plating, selection and analysis of mutant plaques, as well as calculation of mutant frequencies proceeded as for DNA-dependent assays (see sections 3.4.1.3., 3.4.1.4., and 3.4.1.5.). Specific error rates were calculated as previously described, but taking into account that the primer used for reverse transcription hybridized at positions +174 to +191 of the *lacZα* nucleotide sequence. Then, errors were detected at 123 template positions for base substitutions and 178 for frameshifts (38 at runs, and 140 at non-runs). Despite the different target size for base substitutions and frameshifts, we determined overall error rates to facilitate the comparison of different retroviral RTs by considering 200 template positions where any phenotypic change could be detected (see **Figure 14**, except positions +174 to +191). Statistical analysis about the differences in the proportion of mutations at specific sites in the spectra of different RTs was determined by using a two-tailed Fisher's exact test, using the GraphPad software.

3.5. Promoter-dependent transcription assays

The plasmid pTRI-β-actin-Mouse (included in the Ambion MAXIscript® T7 Transcription Kit) was used by T7 RNAP as the promoter-dependent template in transcription assays. These assays were used to assess nucleotide misincorporation efficiencies in the absence of one NTP or with unbalanced mixtures of NTPs, under different transcription conditions. First, the plasmid was digested for 90 min at 37 ºC with EcoRI (New England Biolabs) in 50 µl of 100 mM Tris-HCl buffer pH 7.5, containing 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton X-100 and 3 units of the enzyme per µg of plasmid. The resulting linearized plasmid was analyzed in 1% agarose gels and purified with the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions.

Transcription assays were carried out at 37 ºC for 0-30 min in different buffers depending on the desired pH (i.e. MES, PIPES, Bis-Tris and Tris-HCl). Reactions were done in 20 µl of 40 mM buffer (pH 5.5-7.5), containing 10 mM NaCl, 0.5-6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol (DTT), 5 units of RNasin® Plus RNase inhibitor (Promega), 15 units of T7 RNAP (Ambion) and 0.5 µg of the linearized pTRI-β-actin-Mouse plasmid. Reactions were initiated by the addition of different combinations of NTPs at 100 µM each [e.g., * (ATP, GTP, CTP, UTP); −A (GTP, CTP, UTP); −G (ATP, CTP, UTP); and −U (ATP, GTP, CTP)], as well as with unbalanced mixtures of the four NTPs, including one that is added in a low concentration [e.g., −A↑ (GTP, CTP and UTP at 100 µM each, plus 1 µM ATP); or −G↑↑ (ATP, CTP and UTP at 100 µM each, plus 20 µM GTP)], always in presence of 0.2 µCi of [α-^{32}P]CTP (Perkin Elmer). Reactions were quenched at appropriate times by adding the stop solution [10 mM EDTA in 90% formamide, containing 3 mg/ml xylene cyanol FF and 3 mg/ml bromophenol blue]. The RNA products were resolved on denaturing polyacrylamide gel electrophoresis (20% polyacrylamide and 8 M urea). Then, gels were exposed to a phosphor screen and scanned with a BAS1500 PhosphorImager instrument (Fuji).

3.6. Promoter-independent single-nucleotide transcription assays

The transcription (DNA-dependent RNA synthesis) fidelity of T7 RNAP was studied by single-nucleotide incorporation assays with correct and incorrect nucleotides at pH 7.9 in the presence of 6 mM $MgCl_2$, and at pH 6.75 in the presence of 1.5 mM $MgCl_2$. Due to the many constrains in studying transcription elongation starting from the promoter sequence, we chose to use a promoter-free elongation substrate that allows to bypass the nonprocessive stages of initiation and study the fidelity of RNA synthesis only in the elongation phase. Promoter-free elongation substrates have already been used to dissect the kinetic pathway of T7 RNAP (Anand and Patel, 2006; Tang *et al.*, 2011). Thus, assays were carried out with the VSR10 template-primer (Anand and Patel, 2006), a hybrid containing double-stranded DNA and a short RNA oligonucleotide, lacking the T7 RNAP promoter (**Table 7**).

Table 7. Primers used in promoter-independent transcription assays with T7 RNAP

Oligonucleotides	Sequence
Primer RNA	5′-GGGGGCUGGC-3′
Non-template DNA	5′-GACTACAGACAGGACCCACATAGGATC-3′
Template DNA	5′-CCCGACCGAGCTGATGTCTGTCCTGGGTGTATCCTAG-3′

Template DNA, non-template DNA and primer RNA were mixed at a 1:1.5:1 ratio, heated at 95 °C for 20 min, and then progressively cooled at 75, 55, 45, 20 and 4 °C for 20 min each. The 10-nt primer RNA (3 pmol) was previously labeled at its 5' end with 10 µCi of [γ-^{32}P]ATP (Perkin

Elmer) and 5 units of T4 PNK (New England Biolabs) in 20 µl of 70 mM Tris-HCl buffer pH 7.6, containing 10 mM $MgCl_2$ and 5 mM DTT. The labeling was performed at 37 °C during 45 min, and the PNK was heat-inactivated by incubating the reaction at 90 °C for 10 min. Before annealing, labeled RNA primer was diluted until 100 µl to purify it with a Micro bio-Spin™ column (BioRad) loaded with Sephadex G-25 (GE Healthcare) in order to eliminate the excess of labeled ATP that could interfere with the transcription reaction.

The sample was centrifuged 4 min at 4000 rpm. In order to determine the concentration of the RNA after its purification and to ensure the correct incorporation of the radioactive phosphorous, the samples before and after using the Micro bio-Spin™ columns were chromatographed in PEI-cellulose plates (TLC, 20x20 cm, Merck) using 0.5 M disodium phosphate as mobile phase. The PEI-cellulose sheets were exposed to phosphoroimaging plates (Fujifilm BAS-MP 2040S) and the intensity of both RNA samples was determined with a BAS 1500 scanner (Fujifilm), and using the software TINA 2.09 (Raytest Isotopenmessgerate GmbH, Staubenhardt, Germany). This method was used to estimate the final RNA concentration.

Kinetic parameters for UTP and CTP incorporation were determined at 37 °C using a rapid quench-flow instrument (model QFM-400, Bio-Logic Science Instruments), upgraded with a mixer cross and a special mixer (Bio-Logic). Assays were carried out in 24 µl of 40 mM Tris-HCl buffer pH 7.9, containing 6 mM $MgCl_2$ (or Bis-Tris buffer pH 6.75 with 1.5 mM $MgCl_2$), 10 mM NaCl, 2 mM spermidine and 10 mM DTT, in the presence of increasing concentrations of NTP. All magnesium was loaded in the syringe of the instrument corresponding to the mixture of T7 RNAP and VSR10, facilitating its positioning in the active site before starting the reaction. Due to the short time reactions (ms), this strategy allowed to avoid dramatic reductions in the catalytic constants, especially when low magnesium concentration was used (1.5 mM). The concentration of VSR10 was 240 nM and the recombinant T7 RNAP was added to a final concentration of 1 µM. Excess of enzyme ensured single turnover kinetic conditions for elongation rate measurements. Reactions were stopped with EDTA (0.3 M final concentration).

Pre-steady-state kinetic data were fit to a burst equation: $[P] = A \times [1\text{-}\exp(\text{-}k_{obs} \times t)] + k_{ss} \times t$, where [P] is the product concentration, A is the amplitude of the burst, k_{obs} is the apparent kinetic constant of formation of the phosphodiester bond and k_{ss} is the kinetic constant of the steady-state linear phase. The dependence of k_{obs} on the NTP concentration is described by the Michaelis-Menten equation: $k_{obs} = k_{pol} \times [dNTP]/(K_d + [dNTP])$, where K_d is the equilibrium constant and k_{pol} is the catalytic rate constant of the nucleotide incorporation reaction. Kinetic parameters were determined using curve-fitting tools provided by the SigmaPlot software (Systat Software Inc.).

The incorporation of non-complementary nucleotides GTP and ATP was performed manually with template-primer VSR10 (240 nM). Reactions were carried out in 20 µl of 40 mM Tris-HCl

buffer pH 7.9 (or Bis-Tris buffer pH 6.75), containing 6 mM $MgCl_2$ (or 1.5 mM $MgCl_2$), 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 24 units of RNasin® Plus RNase inhibitor. The T7 RNAP was supplied at 1 µM. Samples were preincubated at 37 °C for 10 min, and the reaction was then initiated by adding increasing concentrations of NTP. Reactions were stopped at different times (0-30 s) by mixing an aliquot of 4 µl with an equal volume of stop solution. Samples were incubated at 90 °C for 10 min before loading in polyacrylamide-urea gels. Results were analyzed by phosphorimaging and nucleotide incorporation data were fitted to the Michaelis-Menten equation, as described above.

3.7. Binding affinity of T7 RNAP for template-primer VSR10

The equilibrium dissociation constant (K_d) for T7 RNAP binding to the template-primer VSR10 was determined after pre-incubating the enzyme with increasing concentrations of the 5'-^{32}P-labeled VSR10 (60 nM-2.3 µM) at 37 °C for 10 min. Then, the incorporation reaction was initiated by adding UTP to a final concentration of 5 µM. Reactions were carried out in 20 µl of 40 mM Tris-HCl buffer pH 7.9 containing 6 mM $MgCl_2$ (or Bis-Tris buffer pH 6.75 containing 1.5 mM $MgCl_2$), and 10 mM NaCl, 10 mM DTT, 2 mM spermidine and 24 units of RNasin® Plus RNase inhibitor. The T7 RNAP was supplied at 0.75-1 µM in reactions carried out at pH 7.9 with 6 mM $MgCl_2$. At pH 6.75 and 1.5 mM $MgCl_2$, the T7 RNAP concentration was 120-250 nM. Aliquots of 4 µl were taken at 10, 20, 30 and 40 s, quenched with stop solution and analyzed by denaturing polyacrylamide gel electrophoresis and phosphorimaging, as described above. The burst amplitudes (RT bound to template-primer at time zero) were plotted as a function of the template-primer concentration, and the data were fitted to the following quadratic equation:

$$[\mathrm{E} \bullet \mathrm{T/P}] = 0.5 \times [(K_d + \mathrm{E_T} + \mathrm{T/P}) - \sqrt{(K_d + \mathrm{E_T} + \mathrm{T/P})^2 - 4 \times \mathrm{E_T} \times \mathrm{T/P}}\,]$$

Where E_T and T/P are the total active enzyme and template-primer concentration used in the assay, respectively, and K_d is the equilibrium dissociation constant for T7 RNAP binding to the template-primer.

3.8. Deep sequencing assays

3.8.1. Reverse transcription

RNA-dependent DNA synthesis accuracy of RTs was determined following the approach of Primer IDs (described in Jabara *et al*., 2011; Zhou *et al*., 2015). Reverse transcription reactions were carried out in 20 µl of 50 mM Tris-HCl buffer pH 8.3, containing 75 mM KCl, 3 mM $MgCl_2$, 100

mM DTT, 1 unit of RNasin® Plus RNase inhibitor (Promega), 500 µM of each dNTP, and 150 nM RT. HIV-1_{BH10} and K65R/V75I_{ESP49} RTs were used in these experiments. Each reaction contained 200,000 RNA template molecules, corresponding to the HIV-1 (LAV) RNA genome. This RNA was purified from supernatants of 8E5 cells (NIH AIDS Reagent Program, Division of AIDS, catalogue number 95) (Folks *et al*., 1986), which contain a single integrated copy of proviral DNA and lacks unintegrated DNA. These cells do not produce RT.

The RNA sample was kindly provided by Drs. Rafael Delgado Vázquez and Mónica García Álvarez (Hospital Doce de Octubre, Madrid, Spain). The primer used to initiate the synthesis of cDNA contained a nucleotide sequence complementary to the PR-coding region of the viral RNA genome (see primer PR-cDNA in **Table 8**). This primer also contained a degenerated twelve-nucleotide sequence, yielding ~16.8×10^6 different barcoded combinations (i.e., 4^{12}). Primers were added to the reaction in excess (62.5 pM), compared to RNA templates. Thus, each cDNA molecule was tagged with an individual and unique barcoded primer.

Table 8. Primers used in deep sequencing experiments to determine the RT's RNA-dependent DNA synthesis fidelity. The HIV-1 (LAV) RNA genome was used as template.

Primers[a]	Sequence (5'-3')	nt[b]
PR-cDNA	GTGACTGGAGTTCAGACGTGTGCTCTTCNNNNNNNNNNNNCAGTTTAACT TTTGGGCCATCCATTCC	67
R-PCR1	GTGACTGGAGTTCAGACGTGTGCTC	25
F-PCR1	CTTCCTACAAGGGAAGGCCAGG	22
R-PCR2	TACGGTAGCAGAGACTTGGTCTGTGACTGGAGTTCAGACGTGTGCTC	47
F-PCR2	ACACTGACGACATGGTTCTACATATAAGAGACAGNNNNCAGGAGCCGAT AGACAAGGAAC	60
R-PCR3	CAAGCAGAAGACGGCATACGAGAT-[BC]-TACGGTAGCAGAGACTTGGTC T	56
F-PCR3	AATGATACGGCGACCACCGAGATCTACACTGACGACATGGTTCTACA	47

[a] F/R, forward/reverse. [b] nt, no. of nucleotides. Sequence in purple indicates the 5' end of the cDNA. The underlined sequences correspond with protease gene sequences. Nucleotide sequences in green and blue correspond to CS2 and CS1 sequences, respectively. Finally, red sequences relate to the Illumina sequencer adaptors.

3.8.2. Amplification and sequencing

The obtained cDNA samples were sent to Fundación Parque Científico de Madrid (FPCM), for further analysis. There, amplicons were generated after three successive PCRs (**Figure 16**), in order to incorporate indexes necessary for sequencing at the termini of the amplicon. Twelve µl of cDNA were amplified in a first PCR with the Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs). The reaction was performed in a total volume of 25 µl of 1X Q5 reaction buffer (New England Biolabs), in the presence of 200 µM of each dNTP, and 500 nM of each primer (R-PCR1

and F-PCR1, see **Table 8**). Samples were heated 2 min at 95 °C, subjected to 30 cycles of 30 s at 95 °C, 1 min at 55 °C and 1 min at 72 °C, and finally incubated for 10 min at 72 °C.

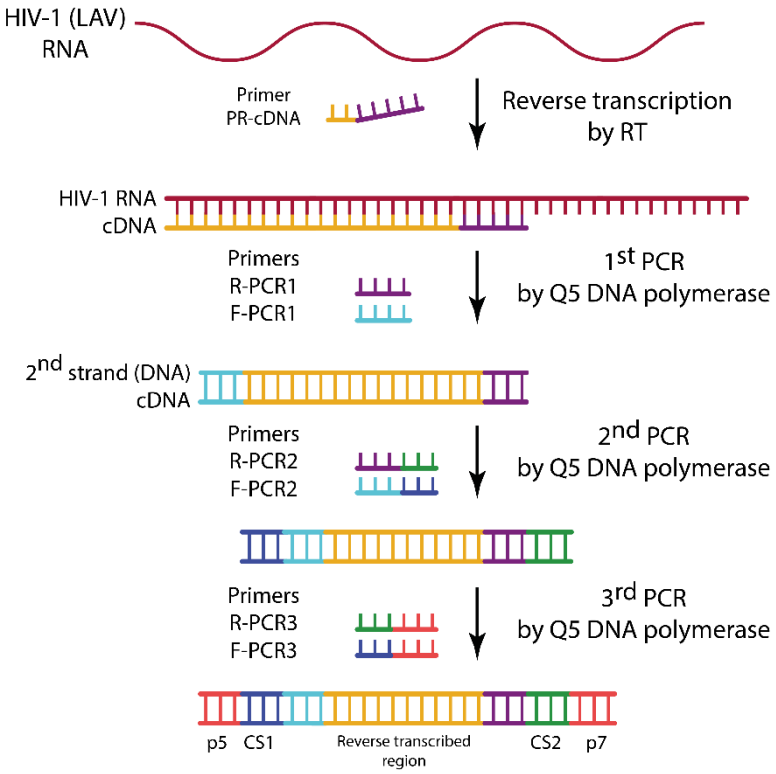


Figure 16. Adaptation of the Primer ID approach to the MiSeq platform. MiSeq library contruction resulting from the reverse transcription of HIV-1 (LAV) RNA genome was used for sequencing. The synthesis of cDNA was initiated with degenerate primers (Primer IDs) containing twelve positions with random nucleotides, providing a barcode for each cDNA molecule synthesized. PR-cDNA primers represent the Primer IDs. First PCR synthesized the second strand of cDNA and amplified the available DNA. Following PCRs introduced indexes needed for the Illumina platform: CS1 (blue) and CS2 (green) during second PCR, and p5 and p7 sequences (red) in the third PCR.

Subsequently, 1 µl of the first PCR product was used as input in the second PCR, which was equivalent to the previous one, except for the primers used in the reaction (R-PCR2 and F-PCR2, see **Table 8**). These primers contain the sequences CS1 and CS2, needed to introduce the sequencer indexes. Then, a third PCR was performed in a total volume of 20 µl of 1X Q5 reaction buffer (New England Biolabs), with the enzyme Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), and in the presence of 400 nM of primers (R-PCR3 and F-PCR3, from the Access Array Barcode Library for Illumina Sequencers –Fluidigm–, see **Table 8**). One µl of the second PCR product was used for the third amplification. This PCR was programmed to carry out the reaction for 30 s at 95 °C, then 14 cycles of 10 s at 98 °C, 20 s at 60 °C, and 20 s at 72 °C, and a final elongation phase of 2 min at 72 °C. This PCR provided the final amplicon containing the indexes required for the Illumina sequencer.

The obtained amplicons were validated and quantified by an Agilent 2100 Bioanalyzer using DNA7500 chips. Equimolecular pools of these amplicons were purified by using agarose gel electrophoresis in order to eliminate primers and dimers found in excess. The purified pools were titrated by quantitative PCR using the "Kapa-SYBR FAST qPCR kit for LightCycler®480", and a reference standard for quantification. Finally, the pool of amplicons at a concentration of 10 pM was denatured prior to be seeded on a flowcell, where clusters were formed and sequenced using a "MiSeq Reagent Kit v3", in a 2×300 pair-end sequencing run on a MiSeq sequencer.

3.8.3. Analysis of sequences

Raw reads were originally provided as FASTQ files, and were analyzed at the Genomics and NGS core facility of the Centro de Biología Molecular Severo Ochoa, which is part of the CEI UAM+CSIC (Madrid, Spain). After performing a quality analysis over reads using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), paired reads were joined to obtain one single sequence. This process was performed with a fast and accurate paired-end read merger called PEAR, which evaluates all possible paired-end overlaps without requiring the target fragment size as input. PEAR also implements a statistical test for minimizing false-positive results.

Then, a cutoff value for the minimum number of cDNA molecules required to generate a consensus was established, taking into account the number of maximum reads obtained for the most represented barcode in the sample. Thus, it would be possible to eliminate those consensus made of barcodes generated as a consequence of a mutation (offspring) in the most represented barcodes, also refers to as birthday problem. The cutoff value was calculated by using the following formula, generated by Zhou and colleagues (2015): $\boldsymbol{c_v = N_0 + 1.96S_{N_0}}$, where c_v indicates the cutoff value; N_0, the maximum frequency of offspring barcodes for a barcode with an abundance of N sequences; S_{No}, the standard deviation of N_0. A cutoff value of around 76 was obtained (with $N_0 \approx 69$, and $S_{No} \approx 3.5$), by setting the maximum limit at N=10,000 sequences per barcode.

Afterwards, consensus sequences were calculated by using an in-house Python (v2.7) script to resolve PCR and sequencing errors. Consensus were created by selecting at each position the base found with maximum frequency between all the cDNA molecules with the same barcode. Only consensus made from the alignment of a number of cDNA molecules comprised in the range 76-10,000 were considered. Then, consensus were subjected to a multiple alignment using MAFFT software (https://mafft.cbrc.jp/alignment/software/), and later a BLAST analysis was performed for the consensus sequences against the LAV reference sequence, obtained from the NCBI database. Variants indicate errors made by eukaryotic RNAP II while synthesizing the RNA molecule used as template, together with those generated by the RT during reverse transcription.

3.8.4. Calculation of error rates

Error rates for HIV-1_{BH10} and K65R/V75I_{ESP49} RTs were calculated by dividing the number of mutations found for a specific type of error (e.g., base substitutions, frameshifts) by the number of consensus sequences analyzed and the number of positions where this type of error can be found. Base substitutions and frameshifts could be detected at 355 template positions, as well as the overall mutations. Frameshifts may occur at 83 homopolymeric sites and 272 heteropolymeric positions (**Figure 17**).

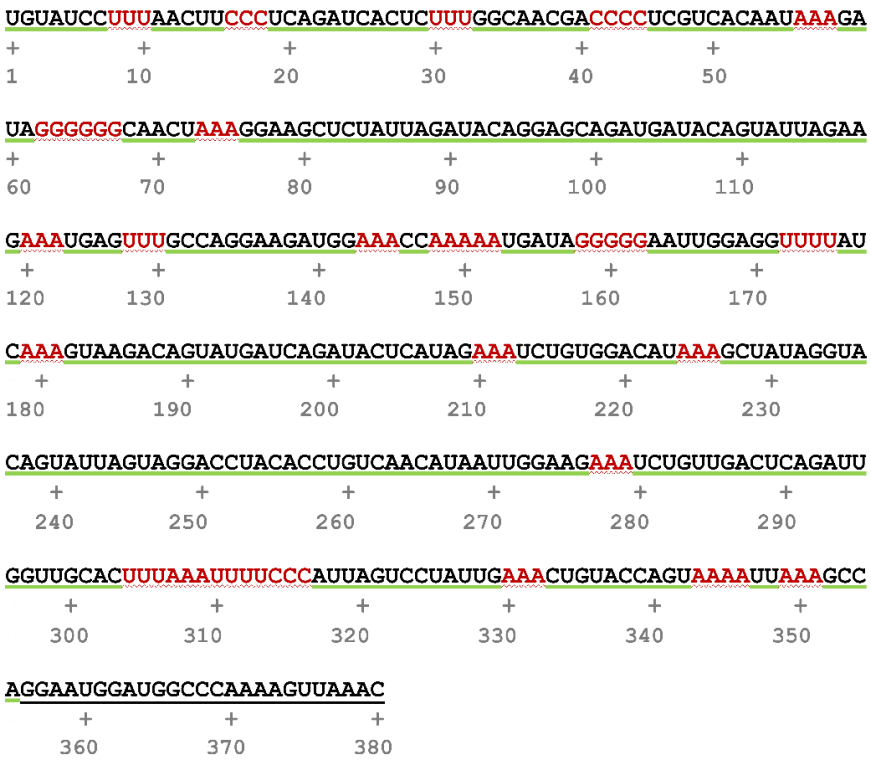


Figure 17. HIV-1 (LAV) RNA sequence targeted in deep sequencing experiments. The primer used to initiate the cDNA synthesis hybridize at positions +356/+380 (sequence underlined in black). Base substitutions and frameshifts are detected at any position between +1 and +355. Sequences underlined in green correspond to heteropolymeric sites, whereas bases colored in red indicate the homopolymeric positions.

4. RESULTS

4.1. Determination of DNA-dependent DNA synthesis fidelity of WT HIV-2_{ROD} RT in comparison with drug-resistant mutants K65R_{ROD} and K65R/Q151M/M184V_{ROD} RTs, and WT HIV-1_{BH10} RT

Nucleoside RT inhibitors (NRTIs) constitute the backbone of current treatments against HIV-1 and HIV-2. Mutational pathways leading to the development of nucleoside analogue resistance are different in both types of HIV. In HIV-2, NRTI-resistance is developed exclusively through the discrimination pathway (i.e., capacity of RT to distinguish between analogue and natural dNTPs), in contrast to HIV-1 that relies either on the exclusion or the excision pathway (Boyer *et al*., 2012). Resistance to all approved NRTIs is conferred in HIV-2 RT by the combination of amino acid substitutions K65R, Q151M and M184V (Smith *et al*., 2009). On the other hand, several amino acid changes in HIV-1 RT leading to drug resistance have been associated with increased fidelity (e.g., M184V) (reviewed in Menéndez-Arias, 2009).

Based on these findings, we analyzed the DNA-dependent DNA synthesis fidelity of resistant RTs K65R_{ROD} and K65R/Q151M/M184V_{ROD}, in comparison with WT HIV-2_{ROD} and HIV-1_{BH10} RTs. In order to determine overall error rates and mutant frequencies for these four RTs, and to analyze rates of specific types of errors (base substitutions and frameshifts), M13mp2 forward mutation assays were carried out. M13-based assays using *lacZα* as a reporter gene provide a broad fidelity assessment, based on a relatively large number of mutational sites and sequence contexts, in contrast to single-nucleotide incorporation assays restricted to a reduced number of substrates, and sequence positions.

By using this method, a dsDNA derived from phage M13mp2, and lacking the *lacZα* gene sequence in one of the two DNA chains (gapped DNA) was used as substrate. Mutations introduced when the RT copied the gapped region (corresponding to the *lacZα* gene) produced a partial or complete loss of activity in the β-galactosidase enzyme. After α-complementation on an appropriate indicator strain, these changes were detected by the altered color phenotype of mutant plaques, when phages were grown in the presence of IPTG and X-Gal. By using this method, silent mutations could not be detected. The method is outlined in **Figure 12**.

Mutant frequencies were calculated as the ratio of mutant (pale blue or colorless) to total plaques. In these assays we found that the WT RT of the HIV-1_{BH10} strain had lower accuracy than the WT HIV-2_{ROD} RT, although differences were relatively small with mutant frequencies of 1.99% and 1.24%, respectively (**Table 9**). Mutations K65R and K65R/Q151M/M184V had almost no effect on the intrinsic accuracy of the HIV-2_{ROD} RT, with mutant frequencies being only slightly lower in comparison with that of the WT enzyme.

Table 9. Accuracy of WT and mutant HIV RTs in M13mp2 *lacZα* forward mutation assays

RTs	Total plaques	Mutant plaques	Mutant frequency[a]
WT HIV-2_{ROD}	9,660	120	0.01242
$K65R_{ROD}$	4,750	56	0.01179 (1.1)
K65R/Q151M/$M184V_{ROD}$	5,142	53	0.01031 (1.2)
WT HIV-1_{BH10}	2,915	58	0.01990 (0.6)

[a]Background frequencies in these assays were estimated to be less than one in 20,000 plaques (Barrioluengo *et al*., 2011). Numbers between parentheses indicate the fold-increase in fidelity relative to the WT HIV-2 RT.

The mutational specificity of the studied RTs was determined after sequencing the *lacZα* mutants generated in the forward mutation assays. Analysis of the obtained mutational spectra revealed a few general trends common to WT and mutant HIV-2 RTs (**Figures 18-21**). For example, more than one third of all errors introduced by HIV-2 RT variants (mutant and WT enzymes) were frameshifts, and in most cases consisted of deletions of one nucleotide.

Two-nucleotide deletions were also present in the spectra of all HIV-2 RTs, representing the 20% of all frameshifts in the case of the WT HIV-2 enzyme. Deletions occurred predominantly at heteropolymeric sequences in the spectra of both HIV-2 and HIV-1 RTs. However, HIV-1_{BH10} RT showed less than 15% of frameshifts in the whole spectrum. Base substitutions were frequently observed in all spectra, with transversions appearing at least three times more frequently than transitions in the analyzed *lacZα* sequence. A predominance of G-to-T substitutions was detected in all mutational spectra, although their frequency was smaller in the case of the K65R/Q151M/$M184V_{ROD}$ RT.

Despite the similarities among the spectra, there were also remarkable differences in the hotspot distribution between WT and mutant HIV-2 RTs (**Table 10**). In our analyses, a hotspot was defined as a position where at least four mutations were observed. The statistical differences in hotspots distribution shown by different RTs were determined by using a two-tailed Fisher's exact test, and are given in **Table 10**. Interestingly, the mutational spectrum of the $K65R_{ROD}$ RT showed one hotspot at position −49, characterized by the presence of one-nucleotide deletions, and another hotspot in which G-to-T transversions accumulate at position +102.

On the other hand, mutations induced by the triple mutant of HIV-2_{ROD} RT clustered at position +82, including various deletions and base substitutions (G-to-T transversions, and G-to-C transitions); and at position +123, where one-nucleotide deletions were detected. All mentioned hotspots were absent in the spectrum of WT HIV-2_{ROD} RT ($P < 0.01$, for comparisons with $K65R_{ROD}$ RT; and $P < 0.03$ for those with the triple mutant RT). It is remarkable that none of the frameshift hotspots described occurred at homopolymeric sequences (or at runs, i.e., those sequences were a base is repeated at least three times).

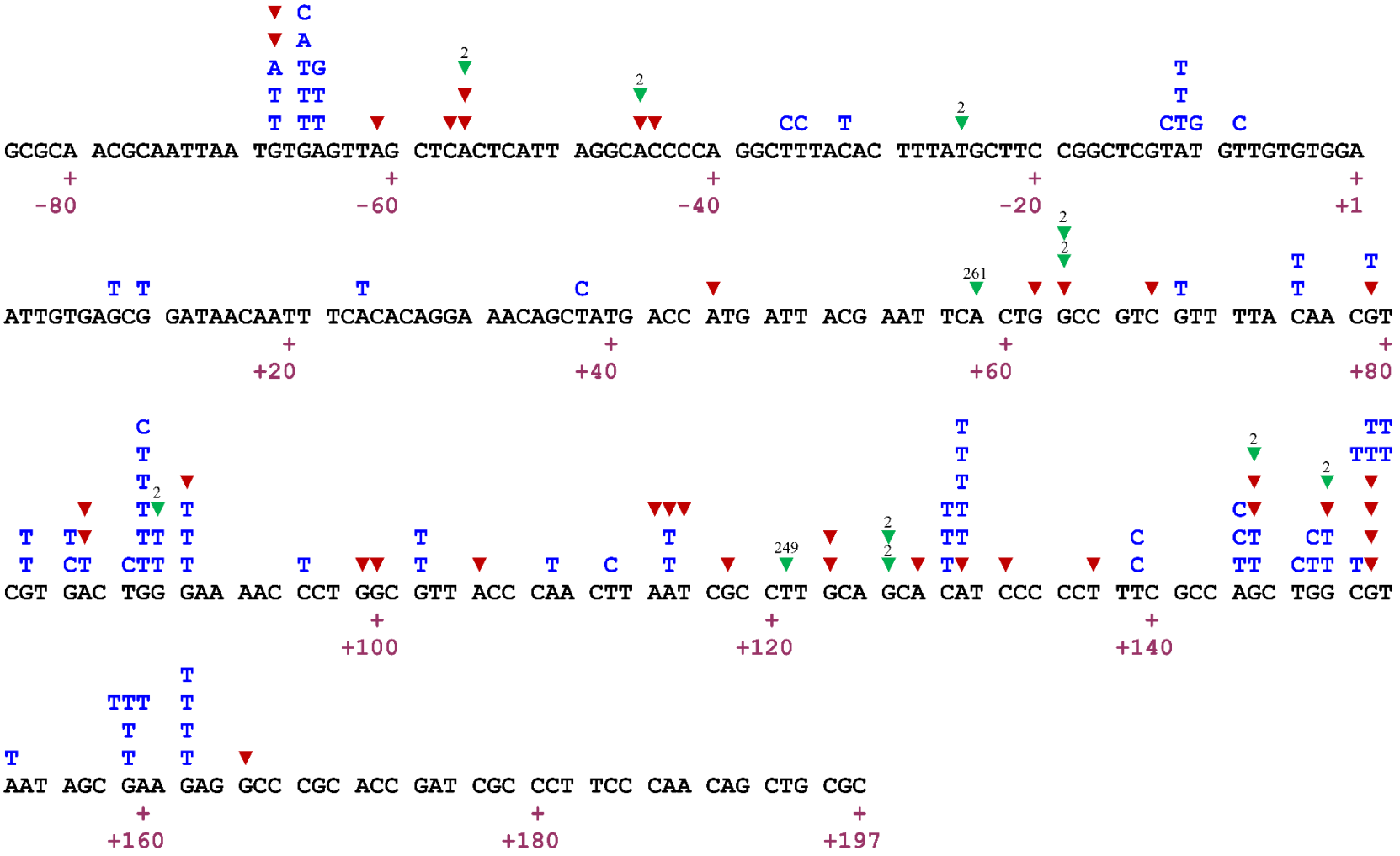


Figure 18. Spectrum of mutations induced by the WT HIV-2$_{ROD}$ RT. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Inverted triangles indicate deletions of one nucleotide (red) or two or more nucleotides (green). On this last case, triangle is positioned in the 3´ end of the deletion and the number of deleted nucleotides is indicated above.

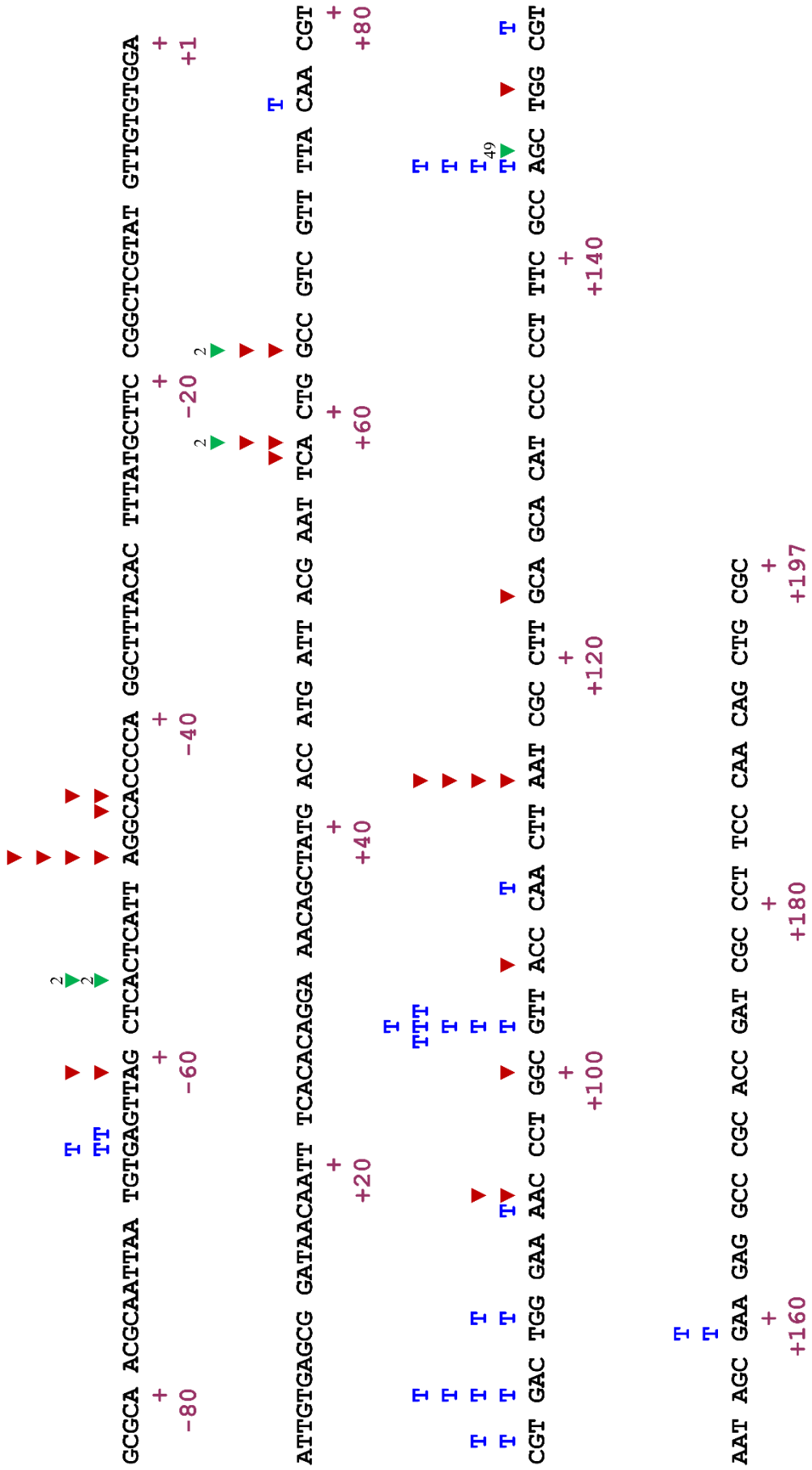


Figure 19. Spectrum of mutations induced by HIV-2 RT mutant K65R. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Inverted triangles indicate deletions of one nucleotide (red) or two or more nucleotides (green). On this last case, triangle is positioned in the 3′ end of the deletion and the number of deleted nucleotides is indicated above.

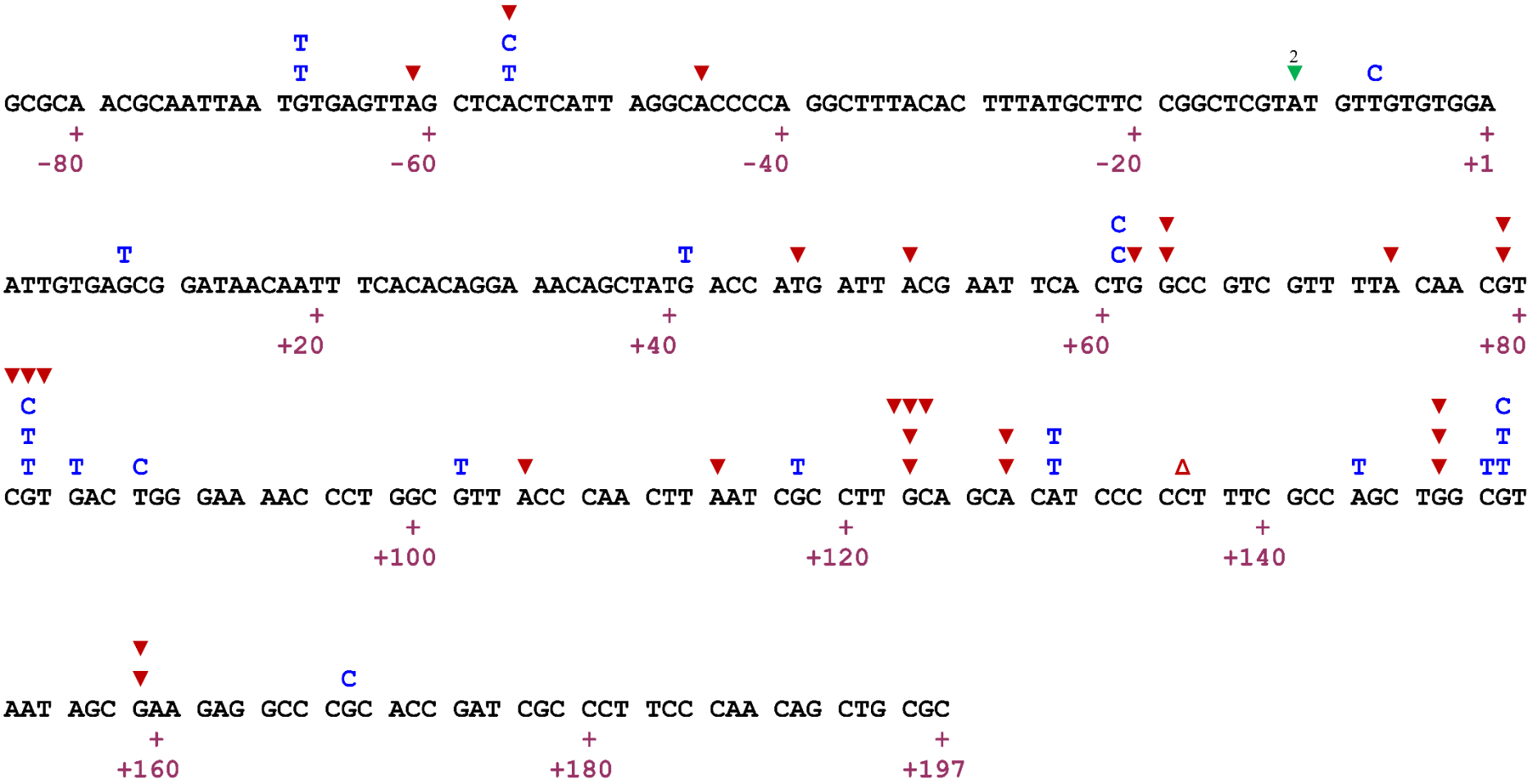


Figure 20. Spectrum of mutations induced by HIV-2$_{ROD}$ RT mutant K65R/Q151M/M184V. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned). Inverted triangles indicate deletions of one nucleotide (red) or two or more nucleotides (green). On this last case, triangle is positioned in the 3´ end of the deletion and the number of deleted nucleotides is indicated above.

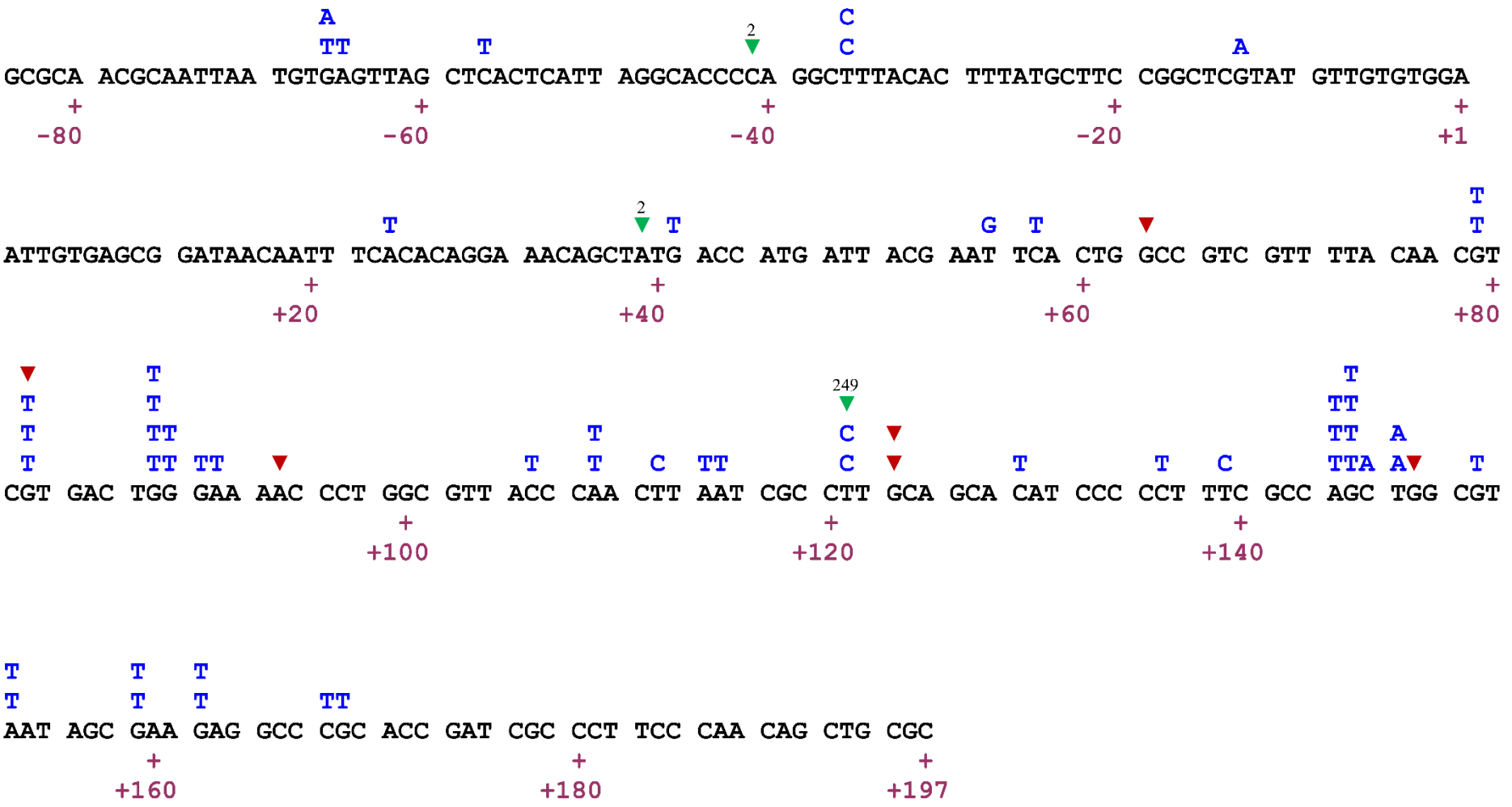


Figure 21. Spectrum of mutations induced by WT HIV-1$_{BH10}$ RT. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lac*Zα target. Inverted triangles indicate deletions of one nucleotide (red) or two or more nucleotides (green). On this last case, triangle is positioned in the 3´ end of the deletion and the number of deleted nucleotides is indicated above.

Table 10. Summary of mutational hotspots for WT and mutant HIV-2_{ROD} RTs, and WT HIV-1_{BH10} in M13mp2 *lacZα* forward mutation assays. Listed hotspots correspond to sites where we found at least four mutations in any of the mutational spectra analyzed.

Mutation type	RTs				Statistical analyses of hotspots [a]
	WT HIV-2_{ROD}	K65R$_{ROD}$	K65R/Q151M/M184V$_{ROD}$	WT HIV-1_{BH10}	
	No. of errors	No. of errors	No. of errors	No. of errors	
All classes	137	57	54	63	
Base substitutions					
−66	5	2	0	2	NSS [b]
+84	2	4	1	0	NSS
+88	6	2	0	4	NSS
+102	2	7	1	0	Comparisons of K65R$_{ROD}$ RT with WT HIV-1_{BH10} and HIV-2_{ROD} RTs render P values of $P<0.02$ and $P<0.005$, respectively
+144	3	4	1	3	NSS
+145	2	0	0	4	NSS
+159	5	2	0	2	NSS
+162	4	0	0	2	NSS
Frameshifts					
−49	0	4	0	0	$P<0.01$ (K65R$_{ROD}$ RT) *versus* HIV-2_{ROD} RT
+114/+115	3	4	1	0	NSS
+123	2	1	5	2	$P<0.03$ (K65R/Q151M/M184V$_{ROD}$ RT) *versus* HIV-2_{ROD} RT
Both					
+82	2	2	6	4	$P<0.02$ (K65R/Q151M/M184V$_{ROD}$ RT) *versus* HIV-2_{ROD} RT

[a] Statistical analyses were carried out for the six possible comparisons between the four RTs, using a two-tailed Fisher's exact test. [b] NSS, Not statistically significant.

[c] Positions −68, +90, +115, +130, +145, +149, +151 also clustered at least four base substitutions or frameshifts (both), but differences between different RTs are not statistically significant.

Table 11. Summary of error rates for WT and mutant HIV RTs, for various classes of mutations, based on M13mp2 DNA-dependent DNA synthesis reactions

Mutation type	WT HIV-2_{ROD} RT		$K65R_{ROD}$ RT		K65R/Q151M/ $M184V_{ROD}$ RT		WT HIV-1_{BH10} RT	
	No. of errors	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**
All classes	**137**	**1/10027**	**57**	**1/11850**	**54**	**1/13541**	**63**	**1/6580**
Base substitutions	**88**	**1/8233**	**28**	**1/12723**	**24**	**1/16069**	**54**	**1/4049**
Transitions	21 (24%)		1 (4%)		5 (21%)		14 (26%)	
Transversions	67 (76%)		27 (96%)		19 (79%)		40 (74%)	
Frameshifts	**49**	**1/17506**	**29**	**1/14545**	**30**	**1/15220**	**9**	**1/28761**
Insertions	0 (0%)		0 (0%)		1 (3%)		0 (0%)	
Deletions	49 (100%)		29 (100%)		29 (97%)		9 (100%)	
At runs[a]	5	1/91577	2	1/112575	1	1/243731	2	1/69086
At non-runs	44	1/9089	27	1/7283	29	1/7341	7	1/17240

[a] A run is considered when there is a row of three or more identical nucleotides.

The analysis of error rates (**Table 11**) demonstrated that all HIV-2 RTs introduced frameshifts with a similar frequency, while differences in fidelity were observed when looking at base substitution error rates. Thus, the frequency of this type of mutations was reduced by less than two-fold in the case of classwide-resistant K65R/Q151M/M184V$_{ROD}$ RT relative to the WT enzyme, while K65R alone produced a mere 1.5-fold increase in nucleotide substitution fidelity.

The comparison of WT HIV-2$_{ROD}$ and HIV-1$_{BH10}$ RTs demonstrated that both enzymes introduced similar types of errors (i.e., frequent one-nucleotide deletions, and more transversions than transitions). However, frameshifts were more abundant in the mutational spectrum of the HIV-2$_{ROD}$ RT (compare spectra shown in **Figures 18 and 21**). In comparison with the WT HIV-1$_{BH10}$ RT, the HIV-2 enzymes were less prone to introduce single nucleotide substitutions, but showed higher overall frameshift error rates due to their strong tendency to introduce deletions at heteropolymeric sequences.

4.2. Determination of RNA-dependent DNA synthesis fidelity of six retroviral RTs

4.2.1. Analysis of mutant frequencies and error rates of WT MLV, AMV, HIV-1_{BH10} and HIV-1_{ESP49} RTs, as well as mutant RTs $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$ using RNA templates

A modified version of M13mp2 *lacZα* forward mutation assays was used to determine mutant frequencies, and error rates of RNA-dependent DNA synthesis for WT MLV, AMV, HIV-1_{BH10} and HIV-1_{ESP49} RTs, as well as mutant RTs $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$. For this purpose, a commercial T7 RNAP was used to synthesize a *lacZα* RNA template that was then reverse transcribed by retroviral RTs (the method is represented in **Figure 15**). The cDNA product was hybridized with a gapped dsDNA lacking the *lacZα* gene in one of the two DNA strands. Errors made by RTs during reverse transcription result in a decrease in α-complementation, and could be detected under appropriate conditions. Mutant frequencies obtained with the six studied RTs, and calculated as the ratio of mutant to total plaques, are given in **Table 12**. These assays showed less than two-fold differences in fidelity between the most accurate and the least faithful RTs.

Table 12. RNA-dependent DNA synthesis fidelity of WT and mutant RTs in M13mp2 *lacZα* forward mutation assays

RTs	Mutant plaques[a]	Total plaques	Mutant frequency[b]
WT HIV-1_{BH10}	52	12,836	0.00405
WT HIV-1_{ESP49}	46	13,347	0.00345 (1.18)
$K65R_{ESP49}$ [c]	54	18,284	0.00295 (1.37)
$K65R/V75I_{ESP49}$	69	23,512	0.00293 (1.38)
WT MLV	36	13,569	0.00265 (1.53)
WT AMV	28	11,250	0.00249 (1.63)

[a] For each enzyme, mutant plaques were obtained after transfection of gapped DNA hybridized with the cDNA product of ten synthesis reactions. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNAP (Promega) in a transcription buffer containing 40 mM Tris-HCl pH 7.9, and 6 mM $MgCl_2$ (full composition given in Materials and Methods). [b] Background frequencies in these assays were estimated to be less than one in 20,000 plaques (Barrioluengo *et al*., 2011). Numbers between parentheses indicate the fold-increase in fidelity relative to the WT HIV-1_{BH10} RT. [c] Results obtained for mutant $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$ RTs, as well as for WT MLV and AMV are based in part on data collected by Verónica Barrioluengo (Barrioluengo, 2013).

The mutational specificity of the studied RTs was determined after sequencing the *lacZα* gene of the mutant plaques generated with the M13mp2-based mutation assays (**Figures 22-27**). All mutational spectra showed a hotspot of one-thymidine insertions at the homopolymeric sequence located at positions +137/+139. As in previous analysis, a hotspot was defined as a position containing at least four clustered mutations, and statistical differences in hotspot distribution between all RTs were determined by using a two-tailed Fisher's exact test (**Table 13**).

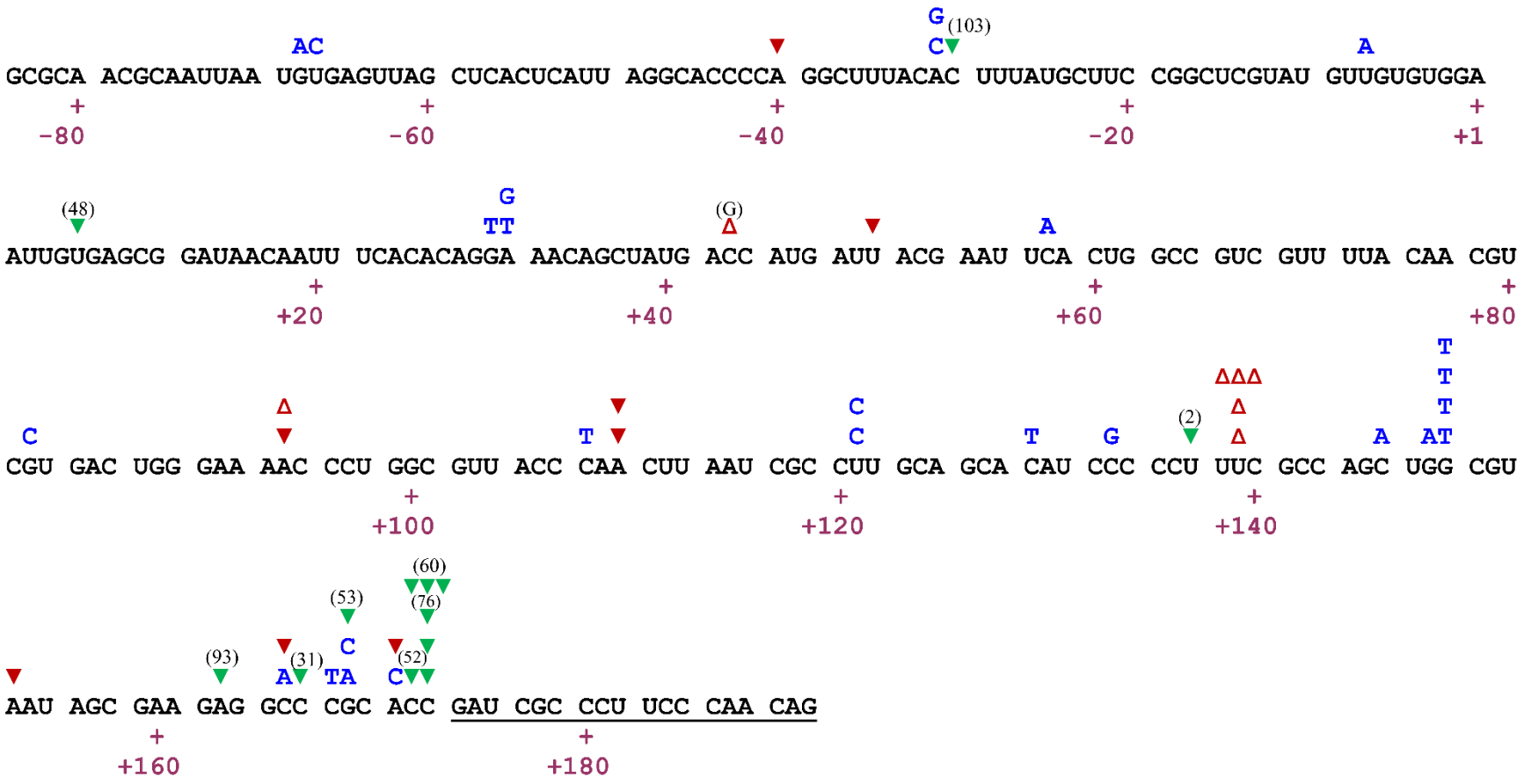


Figure 22. Spectrum of mutations induced by HIV-1$_{BH10}$ RT in M13mp2 *lac*Zα forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lac*Zα target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned, if not indicated). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

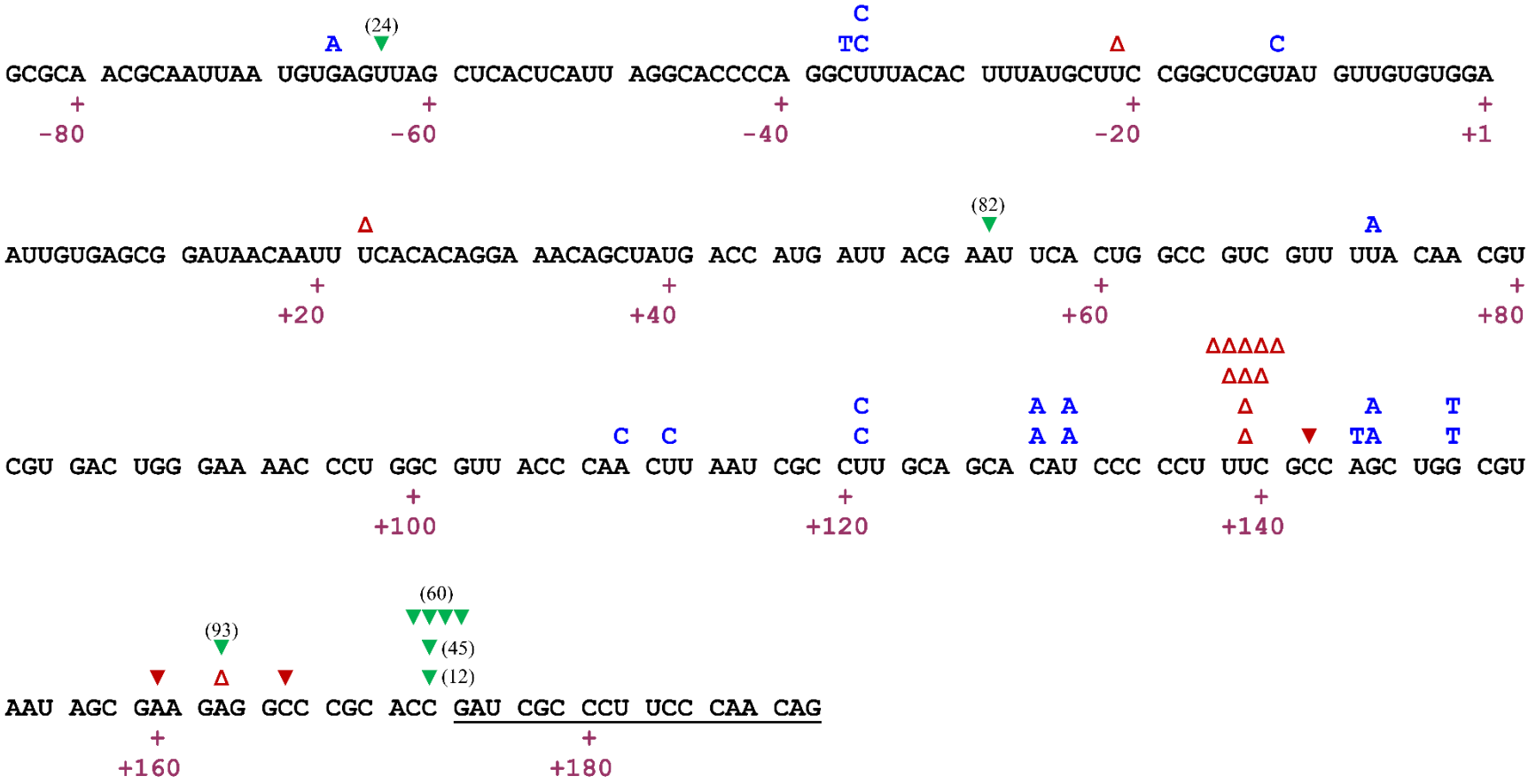


Figure 23. Spectrum of mutations induced by HIV-1_{ESP49} RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lac*Zα target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

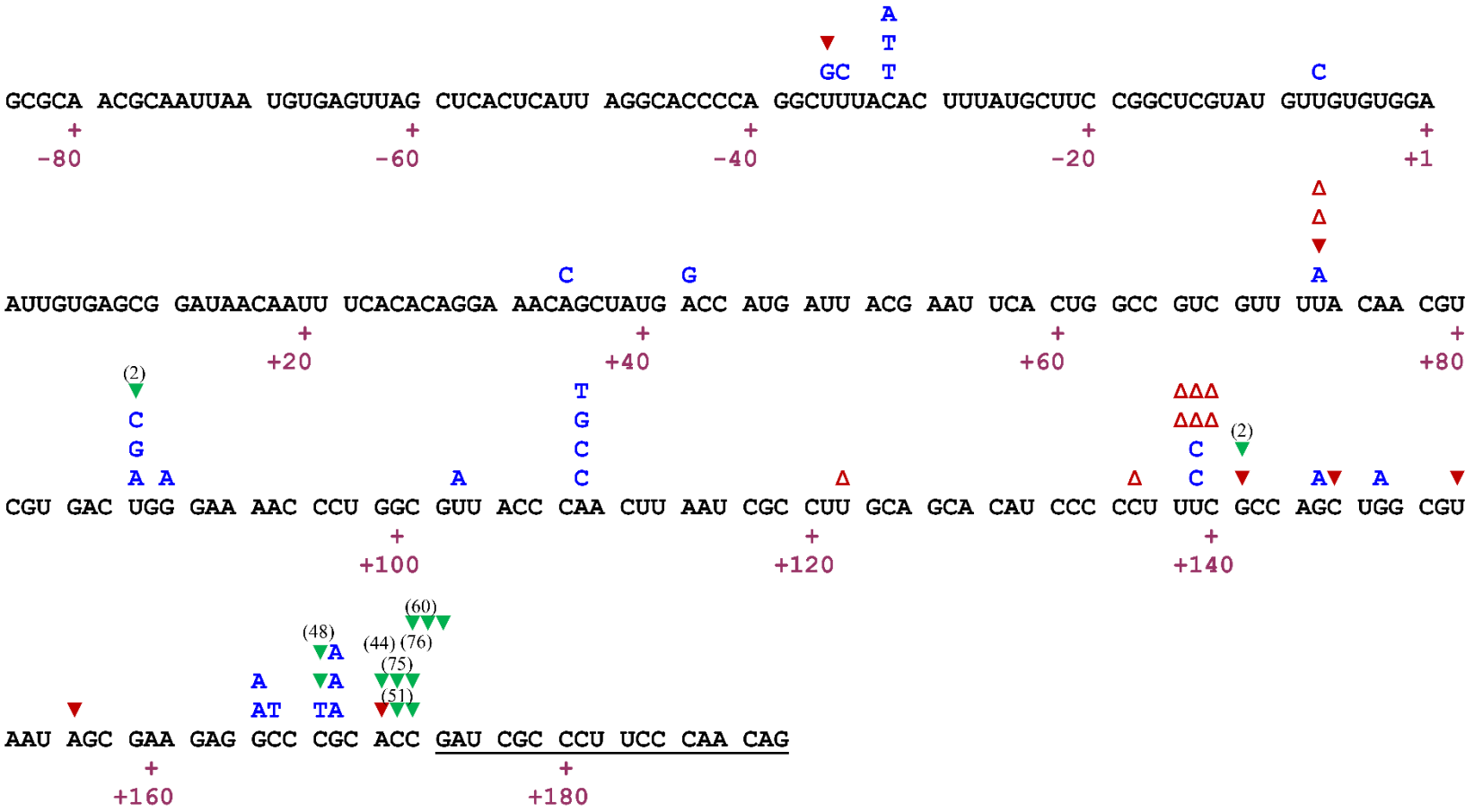


Figure 24. Spectrum of mutations induced by K65R$_{ESP49}$ RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

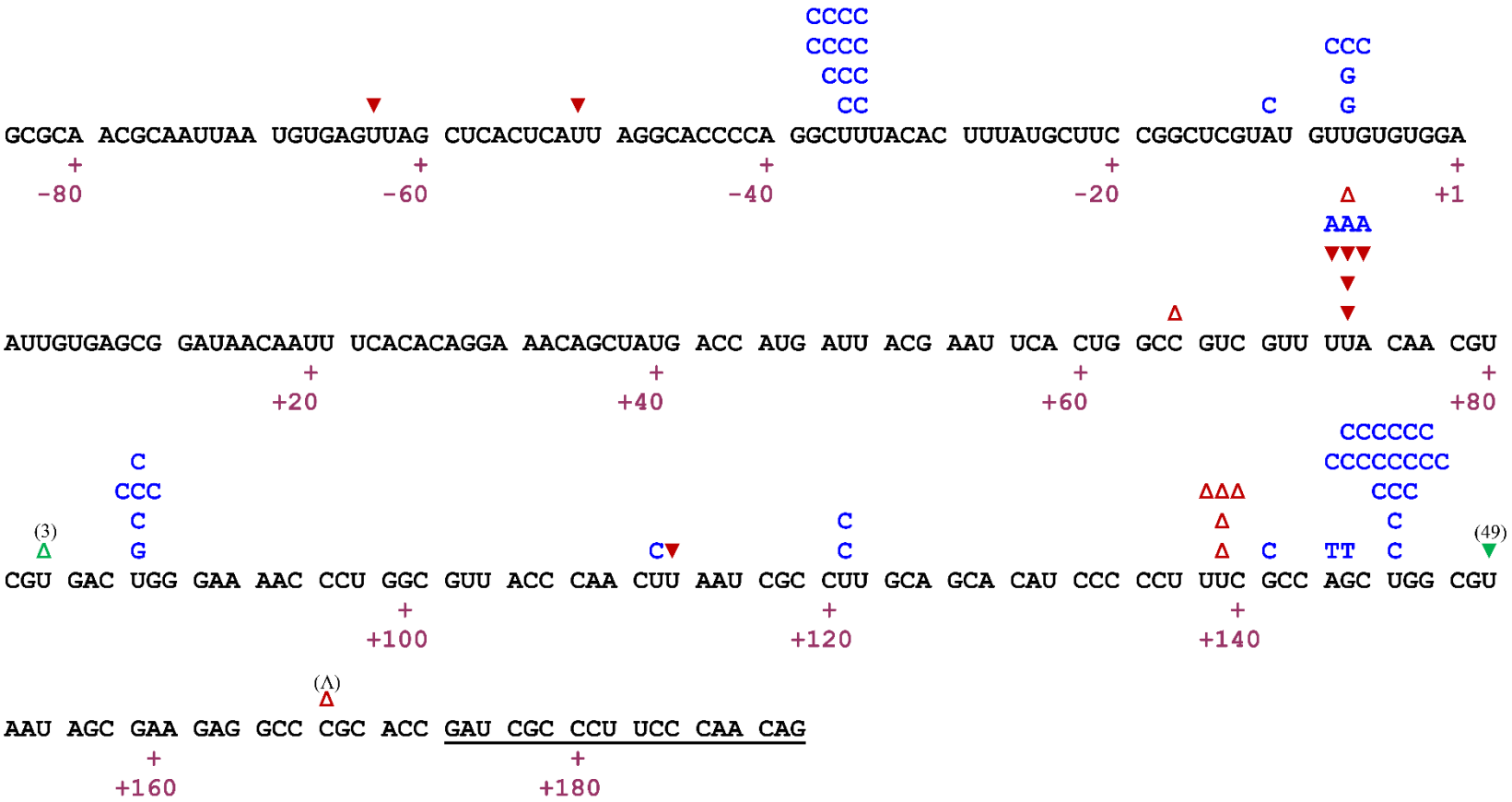


Figure 25. Spectrum of mutations induced by K65R/V75I$_{ESP49}$ RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega)) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright triangles represent insertions of one (red) or more nucleotides (green). In each case, the inserted base (if it is not a duplication of the template base) and the number of inserted bases are indicated in brackets. Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, a triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

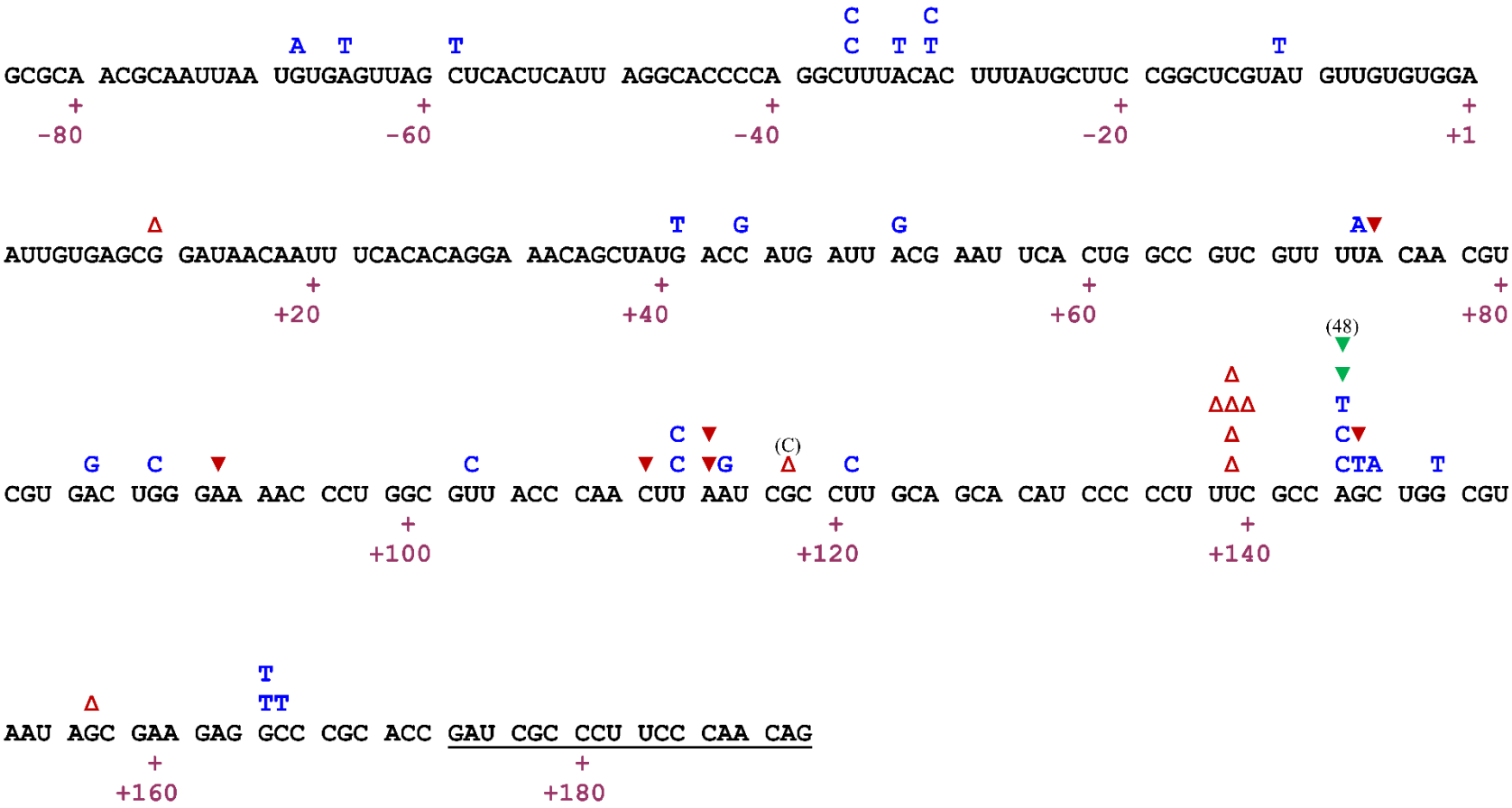


Figure 26. Spectrum of mutations induced by MLV RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned, if not indicated). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, a triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

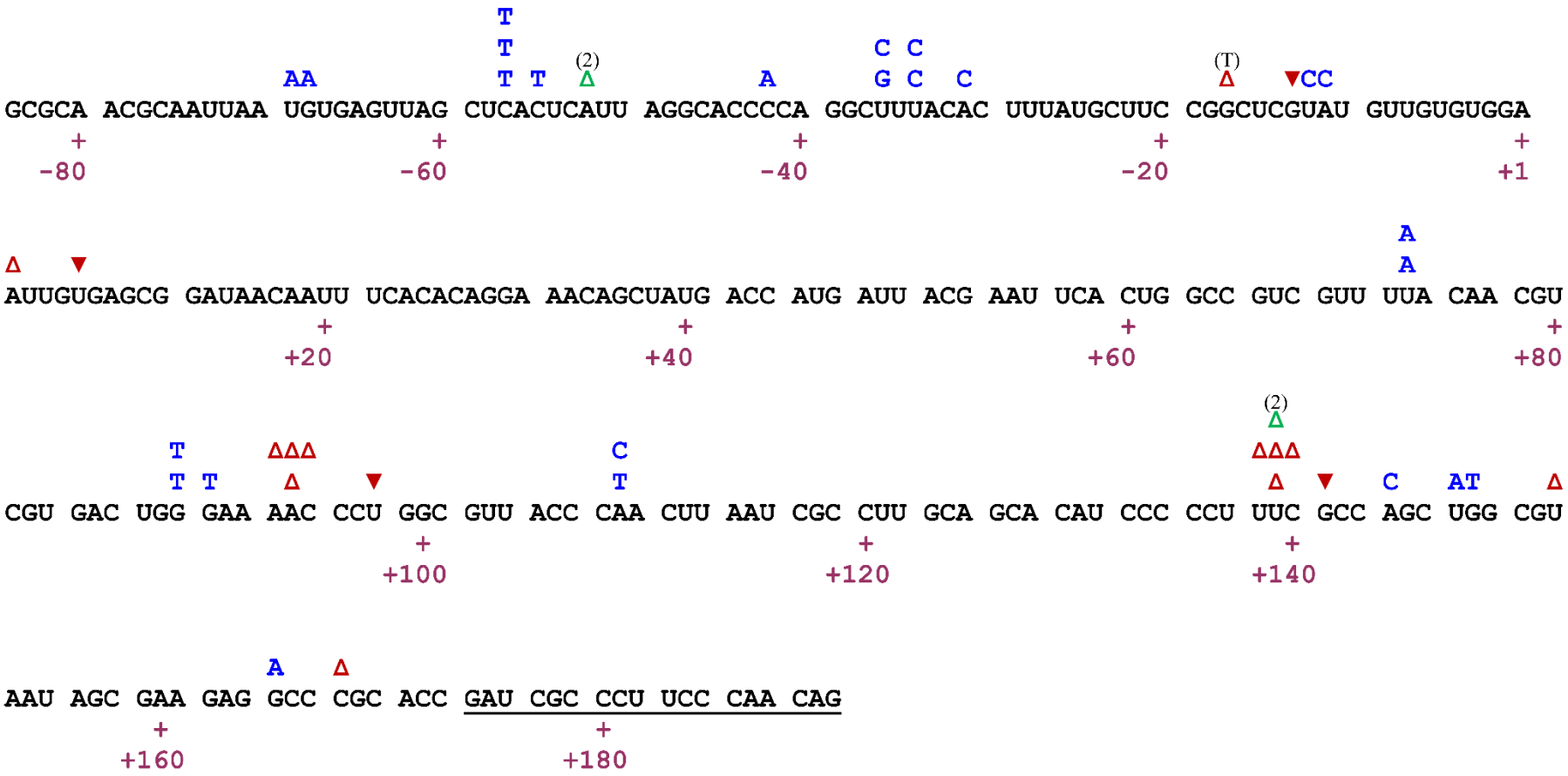


Figure 27. Spectrum of mutations induced by AMV RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright triangles represent insertions of one (red) or more nucleotides (green). In each case, the inserted base (if it is not a duplication of the template base) and the number of inserted bases are indicated in brackets. Inverted triangles indicate deletions of one nucleotide (red). Underlined bases represent the primer used for cDNA synthesis.

Table 13. Summary of mutational hotspots for WT and mutant RTs in M13mp2 *lacZα* forward mutation assays

Mutation type	RTs						Statistical analyses of hotspots [a]
	HIV-1_{BH10}	AMV	MLV	HIV-1_{ESP49}	$K65R_{ESP49}$	$K65R/V75I_{ESP49}$	
	No. of errors	No. of errors	No. of errors	No. of errors	No. of errors	No. of errors	
All classes	54	43	46	44	58	71	
Base substitutions							
-36/-35	0	2	2	2	2	13	$P<0.002$ and $P<0.03$ for the comparison of $K65R/V75I_{ESP49}$ RT *versus* HIV-1_{BH10} and $K65R_{ESP49}$ RTs, respectively
-7	1	0	0	0	1	5	Differences not statistically significant
+87	0	0	0	0	3	6	$P<0.05$ ($K65R/V75I_{ESP49}$ RT) *versus* HIV-1_{BH10} RT
+109	0	2	0	0	4	0	$P<0.05$ ($K65R_{ESP49}$ RT) *versus* $K65R/V75I_{ESP49}$ RT
+147	0	1	0	0	0	19	Strong hotspot in $K65R/V75I_{ESP49}$ RT ($P<0.005$ when compared with all other RTs)
+149	4	0	1	2	0	0	$P<0.05$ (HIV-1_{BH10} RT) *versus* $K65R/V75I_{ESP49}$ RT
Frameshifts							
+94	2	4	1	0	0	0	$P<0.04$ (AMV RT) *versus* $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$ RTs
+139	5	5	6	10	6	5	Differences not statistically significant
+173	6	0	0	6	5	0	Comparisons between HIV-1_{BH10} and HIV-1_{ESP49} RTs *versus* AMV, MLV and $K65R/V75I_{ESP49}$ RTs rendered *P* values in the range of 0.005 and 0.04. $P<0.05$ ($K65R_{ESP49}$ RT) *versus* $K65R/V75I_{ESP49}$ RT. Differences between HIV-1_{BH10}, HIV-1_{ESP49} and $K65R_{ESP49}$ RTs were not significant, as well as differences between AMV, MLV and $K65R/V75I_{ESP49}$ RTs.
Both							
+73	0	2	1	1	4	9	$P<0.02$ ($K65R/V75I_{ESP49}$ RT) *versus* HIV-1_{BH10} RT
+144	0	1	5	1	0	1	$P<0.03$ (MLV RT) *versus* HIV-1_{BH10} and $K65R_{ESP49}$ RTs

The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNAP (Promega) in a transcription buffer containing 40 mM Tris-HCl pH 7.9 and 6 mM $MgCl_2$. Listed hotspots correspond to sites where we found at least four mutations in any of the mutational spectra analyzed.

[a] Statistical analyses were carried out for the fifteen possible comparisons between the six RTs. The analysis was performed using a two-tailed Fisher's exact test.

It is important to highlight that differences in the number of insertions found in the hotspot at positions +137/+139 between the six RTs did not show statistical significance. In addition, U-to-C transitions at positions −36/−35 and U-to-A transversions at position +73 were observed in the spectra of five RTs. On the other hand, we also found large deletions in all mutational spectra, except in the one generated with the AMV RT. In the case of HIV-1_{BH10}, HIV-1_{ESP49} and the $K65R_{ESP49}$ RTs, the large deletions clustered at positions +172/+173 (**Figures 22-24**), corresponding to the first two nucleotides incorporated during reverse transcription. These deletions could derive from misalignment errors occurring while extending the DNA primer.

Despite the similarities found among the mutational spectra, there were also remarkable differences. Thus, for example, AMV RT generated one-nucleotide insertions of adenine in the homopolymeric region located at positions +91/+94 (**Figure 27**), whereas MLV RT showed one hotspot at position +144, including various deletions, and transversions (A-to-C and A-to-T) (**Figure 26**). Interestingly, the spectrum of the K65R/V75I_{ESP49} RT showed an important hotspot at position +147 (U-to-C substitutions) (**Figure 25**). This hotspot is absent in the mutational spectra of the other analyzed RTs (P<0.005, for all five comparisons) (**Table 13**). On the other hand, mutations induced by HIV-1_{ESP49} RT seemed to be scattered throughout the whole *lacZα* sequence (**Figure 23**). The spectrum induced by the single-mutant $K65R_{ESP49}$ had one hotspot at position +109 (**Figure 24**), while in the one obtained with HIV-1_{BH10} RT, G-to-T changes accumulated at position +149 (**Figure 22**).

Error rates for all RTs are summarized in **Table 14**. The highest error rates were obtained with WT HIV-1_{BH10} RT, while one of the most faithful enzymes was the double-mutant K65R/V75I_{ESP49} that showed 1.4-fold increased accuracy relative to the HIV-1_{BH10} RT. Mutant K65R/V75I_{ESP49} RT had a low tendency to introduce frameshifts, although it had a relatively high base substitution error rate. This enzyme was prone to generate transitions (81%), which accumulated at two major hotspots (at positions −36 and +147), and three minor hotspots (located at nucleotides −35, −7 and +87). Unlike in the case of K65R/V75I_{ESP49}, this strong bias towards the generation of transitions was not observed with the other RTs studied. On the other hand, compared with the other RTs, the HIV-1_{BH10} polymerase was prone to introduce frameshift errors, which were predominantly one-nucleotide deletions.

By using the forward mutation assay we cannot determine whether the errors were made by the T7 RNAP during the synthesis of the RNA template (transcription) or by RTs during cDNA synthesis (reverse transcription). The overall error rates obtained with the RTs used in this study were in the range of 2.5×10^{-5} to 3.5×10^{-5}. These values are relatively close to reported estimates of transcription error rates obtained in different organisms (Imashimizu *et al*., 2013; Traverse and Ochman, 2016; Reid-Bayliss and Loeb, 2017; reviewed in Ninio, 1991; Magnuson *et al*., 2016) (see

Table 14. Summary of error rates for WT and mutant RTs, for various classes of mutations in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNAP (Promega) in a transcription buffer containing 40 mM Tris-HCl pH 7.9 and 6 mM $MgCl_2$.

Mutation type	HIV-1_{BH10} RT		AMV RT		MLV RT		HIV-1_{ESP49} RT		K65R $_{ESP49}$ RT		K65R/V75I $_{ESP49}$ RT	
	No. of errors	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**	**No. of errors**	**Error rate**
All classes	**54**	**1/28524**	**43**	**1/31395**	**46**	**1/35397**	**44** [a]	**1/36401**	**58**	**1/37829**	**71**	**1/39739**
Base substitutions	**26**	**1/36434**	**25**	**1/33210**	**29**	**1/34531**	**19**	**1/51843**	**29**	**1/46530**	**53**	**1/32739**
Transitions	11 (42%)		10 (40%)		12 (41%)		10 (53%)		19 (66%)		43 (81%)	
Transversions	15 (58%)		15 (60%)		17 (59%)		9 (47%)		10 (34%)		10 (19%)	
Frameshifts	**28**	**1/48960**	**18**	**1/66750**	**17**	**1/85245**	**25**	**1/57018**	**29**	**1/67336**	**18**	**1/139505**
Insertions	7 (25%)		14 (78%)		9 (53%)		13 (52%)		10 (34%)		9 (50%)	
Deletions	21 (75%)		4 (22%)		8 (47%)		12 (48%)		19 (66%)		9 (50%)	
At runs[b]	10 (36%)	1/29266	10 (56%)	1/25650	7 (41%)	1/44196	12 (48%)	1/25359	13 (45%)	1/32067	12 (67%)	1/44673
At non-runs	18 (64%)	1/59901	8 (44%)	1/118125	10 (59%)	1/113980	13 (52%)	1/86242	16 (55%)	1/95991	6 (33%)	1/329168

[a] The total number of errors with this RT was smaller than the number of mutant plaques (Table 12), because five of those mutant plaques were either recombinants or had complex arrays of mutations, and were not included in the analysis of error rates. [b] A run is considered when there is a row of three or more identical nucleotides.

Discussion), and suggest the existence of an inaccuracy threshold imposed by the T7 RNAP while synthesizing the RNA template.

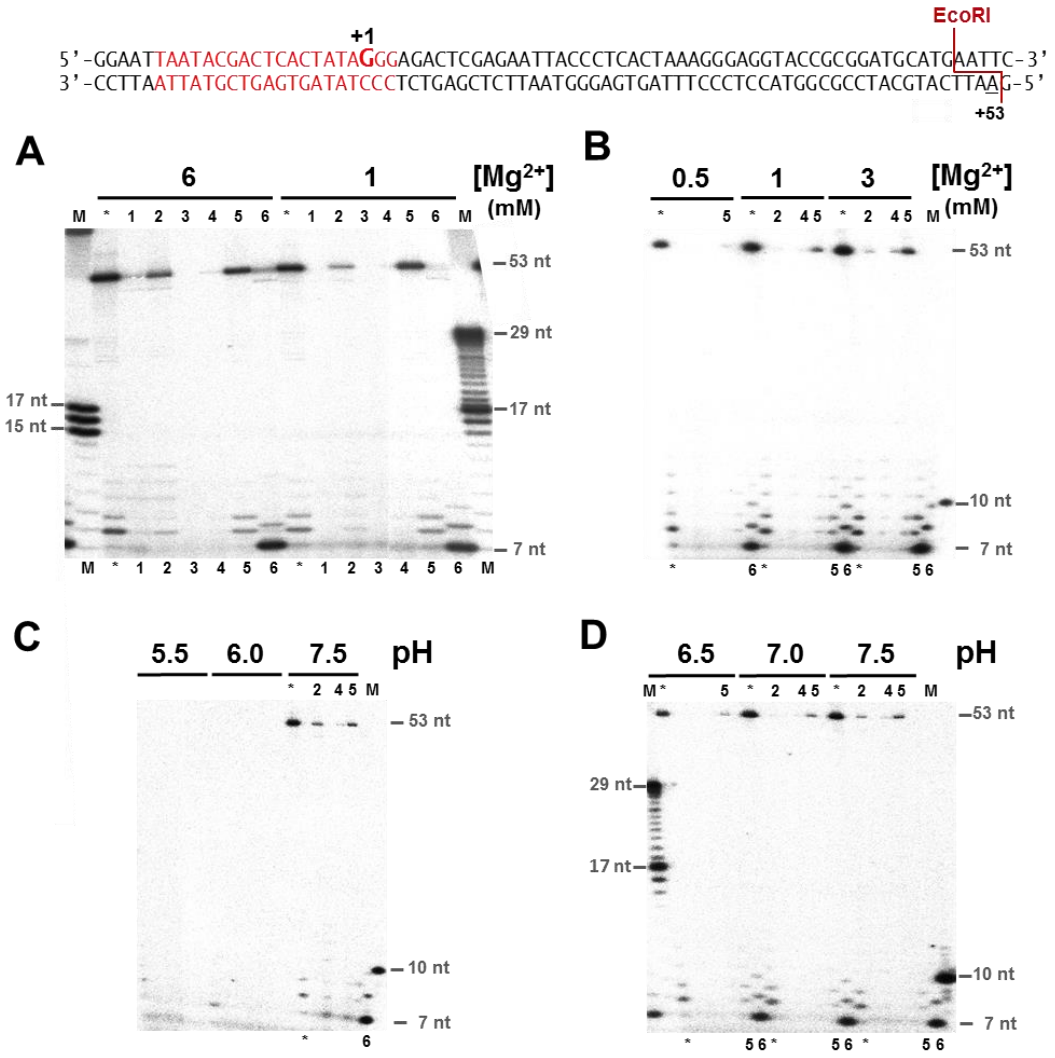
4.2.2. Analysis of bacteriophage T7 RNA polymerase transcription fidelity

4.2.2.1. Fidelity of promoter-dependent transcription under different conditions

In order to reduce the possible inaccuracy threshold imposed by T7 RNAP in M13mp2-based assays, we analyzed its fidelity under different reaction conditions. An EcoRI-linearized pTRI-β-actin-Mouse plasmid containing a T7 promoter sequence was used as template (**Figure 28**). *In vitro* transcription by T7 RNAP (Ambion) was performed in the presence of [α-^{32}P]CTP. The +1 base (G, in bold at the top of **Figure 28**) is the first base incorporated into RNA, as the polymerase transcribes using the opposite strand (3'→ 5') as a template. Reactions were carried out at different pH and Mg^{2+} concentrations, either in the presence of all four NTPs or with biased NTP pools (i.e., lacking or with a very low concentration of one ribonucleotide) (**Figure 28, A-D**). In this assay, when all NTPs were present, the polymerase synthesized an RNA of 53 nucleotides (lanes marked with an asterisk, in **Figure 28**). When using biased pools, the amount of full-length RNA was reduced, since the insertion of incorrect nucleotides would be needed to complete the synthesis reaction. Thus, the full-length RNAs obtained with unbiased NTP pools reflected the inaccuracy of T7 RNAP.

Comparison of relative amounts of full-length products in reactions carried out with biased NTP mixtures and in the presence of all NTPs provided a rough estimate of the accuracy of T7 RNAP in different assay conditions. It should be noted that G is required to initiate transcription, and therefore T7 RNAP cannot synthesize RNA in the absence of this nucleotide (lanes 3 in **Figure 28**). The same happened in the absence of A (lanes 1 in **Figure 28**), except for the reaction carried out in the presence of 6 mM $MgCl_2$, that favoured the incorporation of incorrect nucleotides allowing the visualization of a slight full-length RNA. It is important to indicate that [α-^{32}P]CTP incorporates for the first time in the RNA at position +7. For that reason, reactions performed in the absence of G or A, that were required for initial transcription, did not show intermediate stops. In the case of these two nucleotides, it was essential to perform additional reactions in which G or A were added in a low concentration.

When ATP was supplied at 1 μM while maintaining the three other NTPs at 100 μM (lanes 2), a small but significant amount of full-length RNA was observed at higher magnesium concentrations, as well as above pH 7.0. Similar findings were obtained in reactions with NTP mixtures having low concentrations of GTP (10-20 μM) (lanes 4 and 5 in **Figure 28**). The amount of full-length products in biased reactions was greater at 6 mM $MgCl_2$ and at pH 7.5, suggesting a lower accuracy in those conditions. Inefficient RNA synthesis was also observed in the absence of



Panel	[Mg^{2+}] (mM)	pH	Buffer	Time (min)
A	6	7.5	Tris-HCl	15
	1	7.5	Tris-HCl	15
B	0.5	6.75	PIPES	30
	1	6.75	PIPES	30
	3	6.75	PIPES	30
C	1	5.5	MES	30
	1	6.0	MES	30
	1	7.5	Tris-HCl	30
D	1	6.5	PIPES	30
	1	7.0	PIPES	30
	1	7.5	Tris-HCl	30

Lane	Symbol	rNTP composition[a]
*	*	U + A + G + C at 100 µM each
1	−A	U + G + C at 100 µM each
2	−A^{+}	U + G + C at 100 µM each, plus 1 µM A [b]
3	−G	U + A + C at 100 µM each
4	−G^{+}	U + A + C at 100 µM each, plus 10 µM G
5	−G^{++}	U + A + C at 100 µM each, plus 20 µM G [c]
6	−U	A + G + C at 100 µM each

[a] Panel A (6 mM Mg^{2+}) contained 500 µM each NTP, instead of 100 µM.
[b] Panel A (6 mM Mg^{2+}) contained 2 µM ATP in lane 2.
[c] Panel A contained 50 µM GTP in lane 5.

Figure 28. Promoter-dependent transcription by T7 RNAP in different reaction conditions. Part of the nucleotide sequence of plasmid pTRI-β-actin-Mouse (included in the MAXIscript® T7 Transcription Kit (Ambion)) including the T7 RNAP promoter (in red) and the EcoRI restriction site are shown above. After digestion with EcoRI, the plasmid was used as template for *in vitro* transcription by T7 RNAP in the presence of [α-^{32}P]CTP. The +1 base (G, in bold) is the first base incorporated into RNA during transcription. Then, the polymerase transcribes using the opposite strand (3'→5') as a template. **(A-D)** Panels show transcriptions performed at different concentrations of Mg^{2+} (A and B) and at various pH (C and D). In control reactions (marked *), the four NTPs were supplied at 100 µM, except for reactions shown in panel A (6 mM Mg^{2+}) where their concentrations were increased to 500 µM each. Reactions carried out in the absence of A, G and U are shown in lanes 1, 3 and 6, respectively. Lanes 2, 4 and 5 show primer extensions carried out with limiting amounts of ATP and GTP (indicated in the tables below), and high concentrations of the three other NTPs. M stands for nucleic acid markers.

U (lanes 6 in **Figure 28**), although again more intense bands could be detected with 6 mM $MgCl_2$, like in the absence of A. At pH 5.5-6.0 we did not detect any transcription products even in the presence of all NTPs, indicating that this range of pH is not optimal for T7 RNAP-mediated transcription. In addition, the efficiency of the reaction was very low in the presence of 0.5 mM $MgCl_2$. Taken together, our data indicate that T7 RNAP showed good activity, and improved accuracy at pH 6.5-7.0, and in the presence of Mg^{2+} at 1-3 mM.

4.2.2.2. Nucleotide incorporation kinetics under different conditions

RNA templates used in forward mutation assays were synthesized under optimal conditions for transcription by T7 RNAP (Tunitskaya and Kochetkov, 2002). These reactions were carried out in 40 mM Tris-HCl buffer pH 7.9, containing 6 mM $MgCl_2$ (for further details see Materials and Methods, section 3.4.2.1). However, promoter-dependent transcription assays revealed qualitative differences in fidelity when RNA was synthesized at lower pH, and in the presence of reduced amounts of magnesium. These observations indicated that transcriptions could be more accurate when carried out in buffers at pH 6.75, containing 1.5 mM $MgCl_2$.

The efficiency and nucleotide specificity of T7 RNAP was determined in single-nucleotide incorporation assays using a VSR10 hybrid containing dsDNA and a 10-nucleotide RNA primer, labeled with $\gamma[^{32}P]$ATP (**Figure 29**). These assays were carried out with purified recombinant T7 RNAP due to the requirement of relatively large amounts of enzyme. Nucleotide incorporation rates were determined at different concentrations of correct (UTP) or incorrect ribonucleotides (ATP, CTP and GTP) in standard buffer conditions (40 mM Tris-HCl pH 7.9, containing 6 mM $MgCl_2$), as well as in conditions enhancing T7 RNAP fidelity (40 mM Bis-Tris buffer pH 6.75 and 1.5 mM $MgCl_2$). Kinetic constants for UTP incorporation could not be obtained due to the high efficiency of the polymerization reaction. Thus, at nucleotide concentrations above 300 µM, T7 RNAP showed UTP incorporation rates (k_{obs}) above 150 s^{-1} in both assay conditions. Nevertheless, this is probably an underestimate as the reaction becomes immeasurably fast at increasing nucleotide concentrations. At low concentrations of UTP, the polymerase showed different kinetic behaviours at pH 7.9/6 mM Mg^{2+}, and pH 6.75/1.5 mM Mg^{2+} (**Figure 29A**). The relationship between k_{obs} and [UTP] was linear at pH 7.9/6 mM Mg^{2+} but hyperbolic at pH 6.75/1.5 mM Mg^{2+}.

Similar patterns were observed with the incorporation of CTP. In reactions carried out in 40 mM Tris-HCl buffer pH 7.9, containing 6 mM $MgCl_2$, a linear relationship between k_{obs} and [CTP] was shown while increasing from 0.5 to 4 mM the concentration of nucleotide. In contrast, nucleotide incorporation efficiencies were much lower at pH 6.75/1.5 mM Mg^{2+} and fitted a hyperbola. These data revealed that relative incorporation rates of CTP (incorrect) *versus* UTP

(correct) were lower at pH 6.75/1.5 mM Mg^{2+} than at pH 7.9/6 mM Mg^{2+}. However, reliable estimates of kinetic constants were obtained only for the misincorporation of C at pH 6.75/1.5 mM Mg^{2+} (**Figure 29B**).

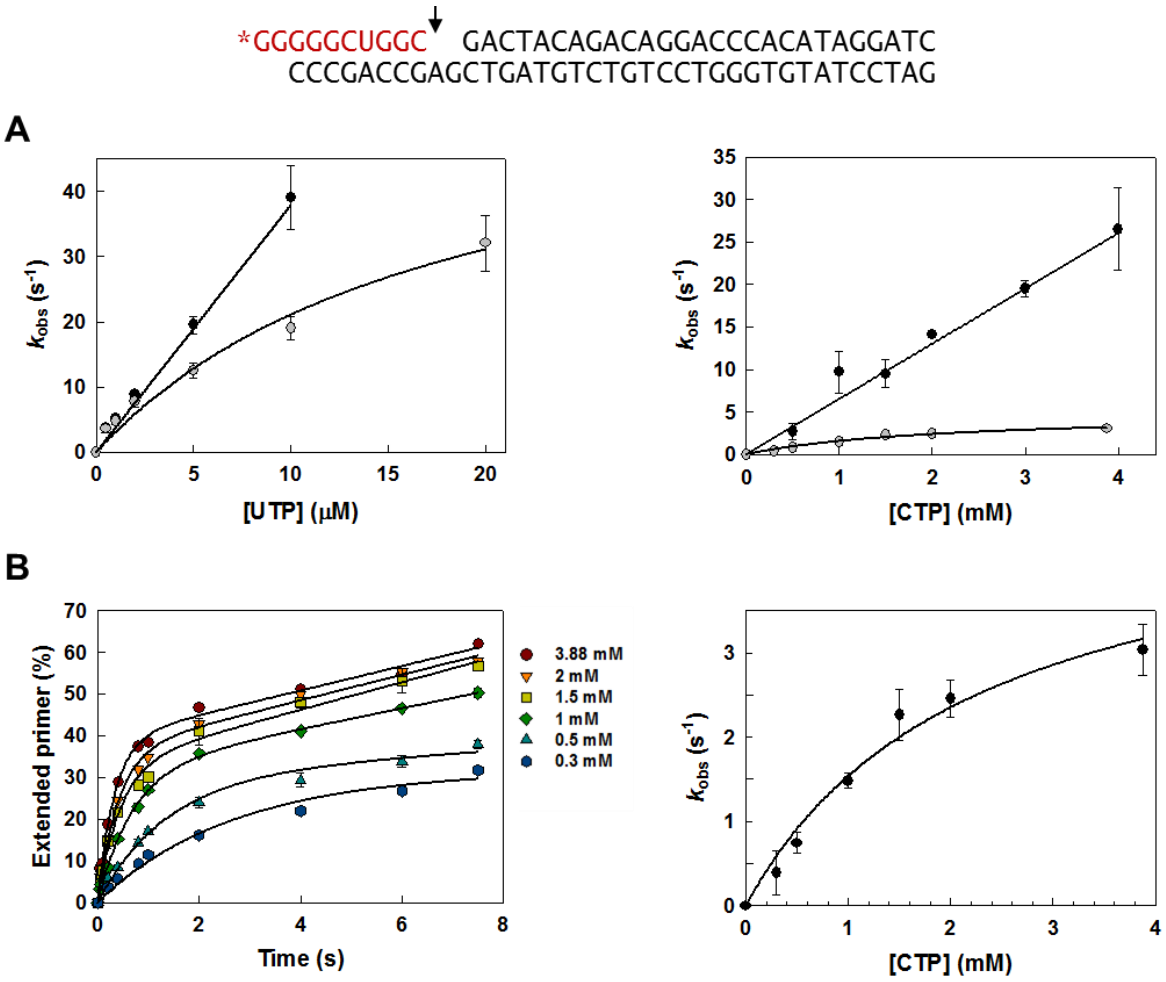


Figure 29. Nucleotide incorporation kinetics for UTP and CTP into VSR10 template-primer by recombinant T7 RNA polymerase. Nucleotide sequences of the VSR10 complex (Anand and Patel, 2006) are shown above. VSR10 is made of a double-stranded DNA (black) annealed to a 10-nt RNA primer (red). The RNA is labeled at its 5' end with [γ-^{32}P]ATP. The arrow indicates the nucleotide incorporation site. **(A)** UTP and CTP incorporation rates at different nucleotide concentrations, obtained at pH 7.9 and 6 mM $MgCl_2$ (black circles), and at pH 6.75 and 1.5 mM $MgCl_2$ (grey circles). **(B)** Pre-steady-state kinetics of CTP incorporation on VSR10 by T7 RNAP, at pH 6.75 (Bis-Tris buffer) and 1.5 mM $MgCl_2$. In the left panel, continuous lines represent the best fit of the data to the single-exponential equation, obtained at different nucleotide concentrations. The right panel shows the nucleotide concentration dependence of CTP incorporation. The continuous line represents the best fit of the k_{obs} data to the Michaelis-Menten equation. The obtained polymerization rates (k_{pol}) and apparent equilibrium dissociation constant (K_d) of CTP were 4.98 ± 0.71 s^{-1} and 2.23 ± 0.61 mM, respectively. Results were obtained from three independent experiments.

Unlike in the case of pyrimidines, the efficiencies of ATP and GTP misincorporation by T7 RNAP were much lower, which allowed their manual determination. This is consistent with the poor incorporation observed in the qualitative assays in the absence of ATP and GTP. Kinetic parameters,

given in **Table 15**, showed that T7 RNAP incorporated non-complementary purines with 2.6- to 3.0-fold higher efficiencies at pH 7.9/6 mM Mg^{2+} than at pH 6.75/1.5 mM Mg^{2+}. ATP was incorporated more efficiently than GTP in both reaction conditions. The catalytic efficiency of ATP incorporation was 29.15 ± 15.17 M^{-1} s^{-1} at pH 7.9/6 mM Mg^{2+} and 11.05 ± 1.53 M^{-1} s^{-1} at pH 6.75/1.5 mM Mg^{2+}. Similar differences were obtained for GTP, although at pH 6.75/1.5 mM Mg^{2+} the k_{pol}/K_d of T7 RNAP was only 6.60 ± 1.16 M^{-1} s^{-1}, while at pH 7.9/6 mM Mg^{2+} was 19.77 ± 10.53 M^{-1} s^{-1}. Taken together, these data provide further support to the notion that at pH 6.75 and 1.5 mM Mg^{2+} the T7 RNAP showed a modest but consistent improvement in its fidelity of RNA synthesis.

Table 15. Nucleotide incorporation kinetics for GTP and ATP into VSR10 template-primer for recombinant T7 RNA polymerase

NTP	pH 7.9 and 6 mM $MgCl_2$			pH 6.75 and 1.5 mM $MgCl_2$		
	k_{pol} ($\times 10^{-2}$ s^{-1})	K_d (µM)	k_{pol}/K_d (M^{-1} s^{-1})	k_{pol} ($\times 10^{-2}$ s^{-1})	K_d (µM)	k_{pol}/K_d (M^{-1} s^{-1})
GTP	1.8 ± 0.6	901.2 ± 334.9	19.77 ± 10.15	4.3 ± 0.7	652.5 ± 50.9	6.60 ± 1.16
ATP	1.7 ± 0.2	583.0 ± 296.6	29.15 ± 15.17	4.1 ± 0.5	367.8 ± 20.5	11.05 ± 1.53

Incorporation rates (k_{pol}) were determined by assuming the concentration of catalytically competent T7 RNAP in the assay (i.e., polymerase bound to template-primer at time zero) determined by using the dissociation equilibrium constants given in **Figure 30**. Data shown are the mean values ± standard deviations obtained from at least three experiments.

4.2.2.3. Template-primer binding affinity of T7 RNAP under different conditions

In addition to its effects on fidelity, we also measured the impact of transcription conditions on the affinity of T7 RNAP for the template-primer. For that purpose, the enzyme was incubated with increasing concentrations of VSR10 complex and the reaction initiated by adding the correct nucleotide (UTP). Interestingly, the assay conditions had a strong influence on template-primer binding. At pH 7.9 and 6 mM Mg^{2+}, the dissociation equilibrium constant (K_d) for T7 RNAP and the VSR10 heteroduplex was about 30 times higher than at pH 6.75 and 1.5 mM $MgCl_2$ (**Figure 30**). The higher affinity for VSR10 of the T7 RNAP at lower pH and Mg^{2+} concentration could represent an additional advantage towards increasing the efficiency of RNA synthesis in the presence of limiting amounts of template.

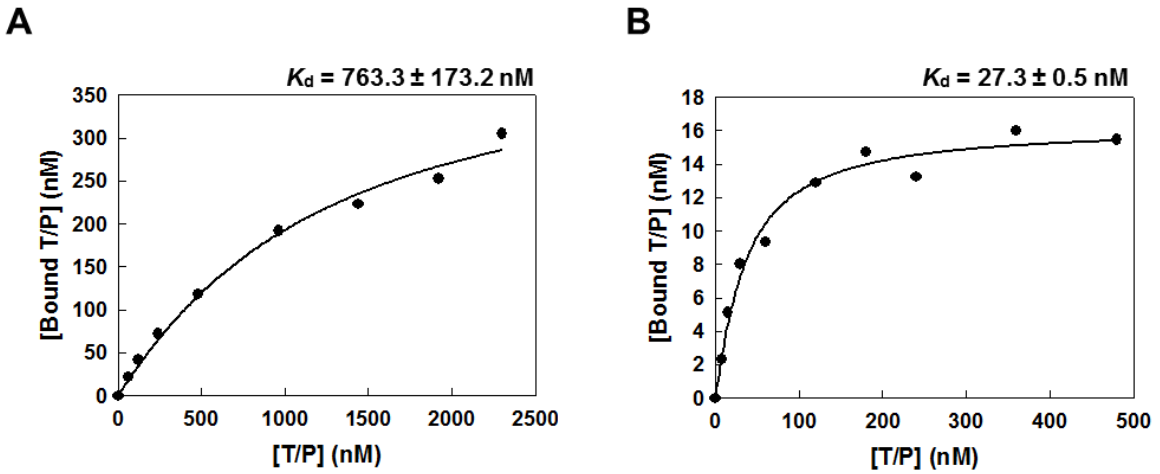


Figure 30. Dissociation equilibrium constants (K_d) for T7 RNA polymerase and the VSR10 dsDNA/RNA heteroduplex. (A-B) Data shown were obtained from representative assays carried out at pH 7.9 and 6 mM $MgCl_2$ (panel A) and at pH 6.75 and 1.5 mM $MgCl_2$ (panel B). The solid line represents the best fit of the data to the quadratic equation relating the template-primer bound to the T7 RNAP and the total concentration of the VSR10 heteroduplex. Reported K_d values represent averages ± standard deviations obtained from at least three independent experiments.

4.2.3. Estimates of RNA-dependent DNA synthesis fidelity of HIV-1 RTs using more faithful RNA templates

Although the biochemical studies revealed that it is possible to increase transcription fidelity by reducing the pH and magnesium concentration of the reaction, it is not clear whether this would affect estimates of fidelity of RNA-dependent DNA synthesis of retroviral RTs. As shown in **Tables 12 and 14**, the highest mutant frequency and overall error rate obtained with RNA templates synthesized at pH 7.9 and in the presence of 6 mM $MgCl_2$ were found in assays carried out with the HIV-1_{BH10} RT.

We used this enzyme to test the impact of the RNA template on mutation frequencies obtained with M13mp2 *lacZα* forward mutation assays. Theoretically, more faithful RNAs (i.e., those carrying less transcription errors) would reduce the mutational threshold of the assay. RNAs containing the *lacZα* sequence were synthesized using recombinant T7 RNAP in the two different conditions (i.e., pH 7.9/6 mM Mg^{2+}, and pH 6.75/1.5 mM Mg^{2+}). Lower mutation frequencies were obtained with HIV-1_{BH10} RT using RNA templates synthesized at lower pH and Mg^{2+} concentrations, although the differences were not large (**Table 16**).

The analysis of the mutational spectra showed similar distributions of transitions, and transversions using both RNAs (**Figures 31 and 32**). The mutational spectra of HIV-1_{BH10} obtained with RNA synthesized at pH 7.9 and in the presence of 6 mM Mg^{2+} accumulate U-to-C transitions at position +121. This hotspot is not present in the spectra obtained with RNA made at pH 6.75/1.5 mM Mg^{2+} ($P < 0.03$ in a two-tailed Fisher's exact test). At the lower pH and Mg^{2+} concentration, we

observed an increased tendency to introduce frameshift errors in homopolymeric *versus* heteropolymeric sequences, but in both cases at a lower rate compared to the spectra obtained at pH 7.9/6 mM Mg^{2+} (**Table 17**). In any case, base substitution and frameshift error rates were about 49-58% lower in assays carried out with RNA obtained using the pH 6.75/1.5 mM Mg^{2+} conditions.

Table 16. RNA-dependent DNA synthesis fidelity of HIV-1 RTs in M13mp2 *lacZα* forward mutation assays

RTs	T7 RNAP source	Transcription conditions		Mutant plaques	Total plaques	Mutant frequency
		pH	[Mg^{2+}] (mM)			
WT HIV-1_{BH10}	Recombinant	7.9	6	45	4,650	0.00968
WT HIV-1_{BH10}	Recombinant	6.75	1.5	70	11,226	0.00624
K65R/V75I_{ESP49}	Recombinant	6.75	1.5	56	14,735	0.00380
K65R/V75I_{ESP49}	Commercial	6.75[a]	1.5	51	19,466	0.00262

Template RNA used in these assays was obtained in T7 RNAP-catalyzed reactions carried out at different pH and [Mg^{2+}], as indicated.

[a] This assay was carried out with the T7 RNAP from Promega in PIPES buffer (pH 6.75), and results were obtained after pooling the products of six reverse transcription reactions.

In previous studies carried out in the laboratory, we found that the mutant K65R/V75I_{ESP49} RT was >15-fold more accurate than the HIV-1_{BH10} RT in forward mutation assays measuring fidelity of DNA-dependent DNA synthesis (Barrioluengo *et al*., 2011). We compared the difference in fidelity of both RTs using M13-based assays with RNA obtained at lower pH and $MgCl_2$ concentration. Although mutant frequencies obtained with the double-mutant RT were lower than those obtained with HIV-1_{BH10} RT using RNAs synthesized at pH 6.75/1.5 mM Mg^{2+} (**Table 14**), the effects were not large. An overall error rate of 3.3×10^{-5} was estimated for the K65R/V75I_{ESP49} RT under those conditions (**Table 17**).

Interestingly, the analysis of mutational spectra revealed that major hotspots obtained with K65R/V75I_{ESP49} RT while reverse transcribing an RNA template synthesized at pH 7.9/6 mM Mg^{2+} (**Figure 25**) were absent from the spectra obtained with the RNA made at pH 6.75/1.5 mM Mg^{2+} (**Figure 33**). Examples of hotspots that were lost with RNAs obtained at lower pH and Mg^{2+} concentration included clusters of U-to-C transitions found at positions −36/−35 and +147 (P < 0.002, for both comparisons), as well as accumulations of U-to-A transversions found at position +73, together with one-nucleotide frameshifts (P < 0.02). It should be pointed out that for the same RT (K65R/V75I_{ESP49}), the number of one-thymidine insertions located in the homopolymeric region +137/+139 was statistically different comparing assays in which RNA was obtained under different transcription conditions (low/high pH and magnesium concentration, and commercial/recombinant T7 RNAP) (P < 0.05).

The analysis of data shown in **Tables 12, 14, 16 and 17** also revealed that RNA synthesized by recombinant T7 RNAP had lower quality than the RNA made with commercial preparations of the enzyme that rendered lower mutant frequencies in M13mp2-based assays, for both HIV-1_{BH10} and K65R/V75I_{ESP49} RTs. This difference in the quality of the RNA is observed even when recombinant T7 RNAP transcribes at pH 6.75/1.5 mM Mg^{2+} and the commercial enzyme at pH 7.9 and high Mg^{2+} concentration. Thus, error rates obtained with the HIV-1_{BH10} RT using RNA synthesized at pH 7.9/6 mM Mg^{2+} were estimated to be 8.4×10^{-5} with the recombinant T7 RNAP, and 3.5×10^{-5} with the commercial enzyme, a difference of 2.4-fold. Similar experiments with K65R/V75I_{ESP49} RT and RNAs transcribed at pH 6.75/1.5 mM Mg^{2+} showed error rates of 3.3×10^{-5} and 2.3×10^{-5} for recombinant and commercial T7 RNAPs, respectively.

The smaller differences observed with the K65R/V75I_{ESP49} RT relative to the HIV-1_{BH10} RT can be attributed to the existence of a transcriptional threshold that has a bigger influence on more faithful RTs. This limit seems to be achieved with RNAs obtained at pH 6.75/1.5 mM Mg^{2+} with the commercial T7 RNAP. However, when the commercial enzyme is used, the pH and Mg^{2+} concentration had a relatively small influence in the error rate (2.5×10^{-5}, with RNAs obtained at pH 7.9/6 mM Mg^{2+}) while differences may not be significant. Interestingly, and in agreement with our proposal, the error rate of HIV-1_{BH10} RT was reduced by 1.6-fold when the RNAs were synthesized by recombinant T7 RNAP at pH 6.75/1.5 mM Mg^{2+}, as compared with those obtained at pH 7.9/6 mM Mg^{2+}. Meanwhile, the smaller reduction in the error rate (1.1-fold) observed with K65R/V75I_{ESP49} RT when changing transcription conditions might be attributed to the use of commercial T7 RNAP to synthesize the RNA templates.

The origin of the T7 RNAP had a minor effect on hotspot distribution in the K65R/V75I_{ESP49} spectra when the template RNA was synthesized at lower pH and magnesium concentration (**Figures 33 and 34**). Thus, major hotspots at positions +87 and +139 were found with RNAs synthesized at pH 6.75/1.5 mM Mg^{2+}, with recombinant or commercial T7 RNAP. Differences in hotspot distribution along the *lacZα* sequence between both assays were not significant statistically.

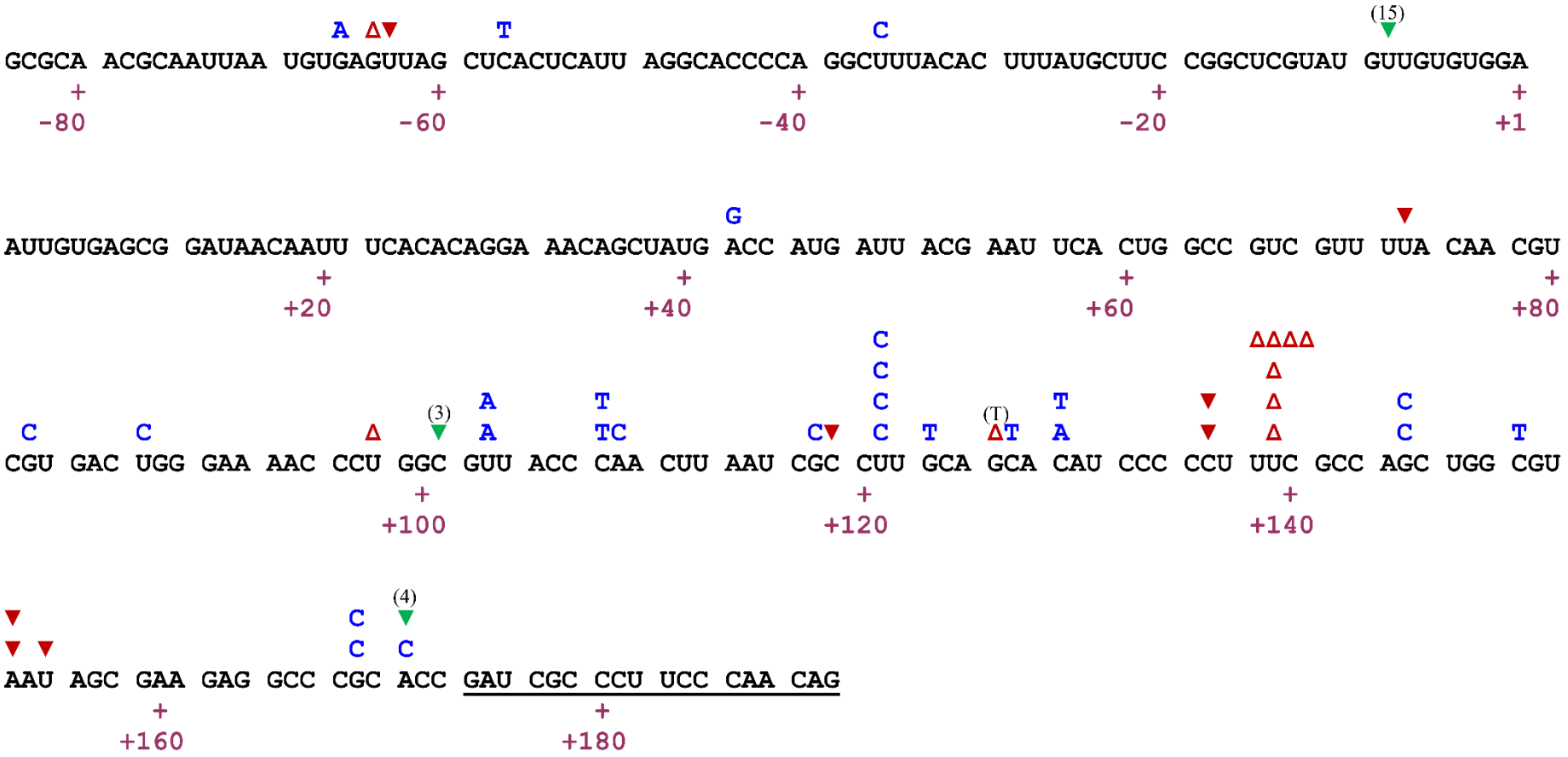


Figure 31. Spectrum of mutations induced by HIV-1_{BH10} RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the recombinant T7 RNA polymerase at pH 7.9 (Tris-HCl buffer) and in the presence of 6 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned, if not indicated). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

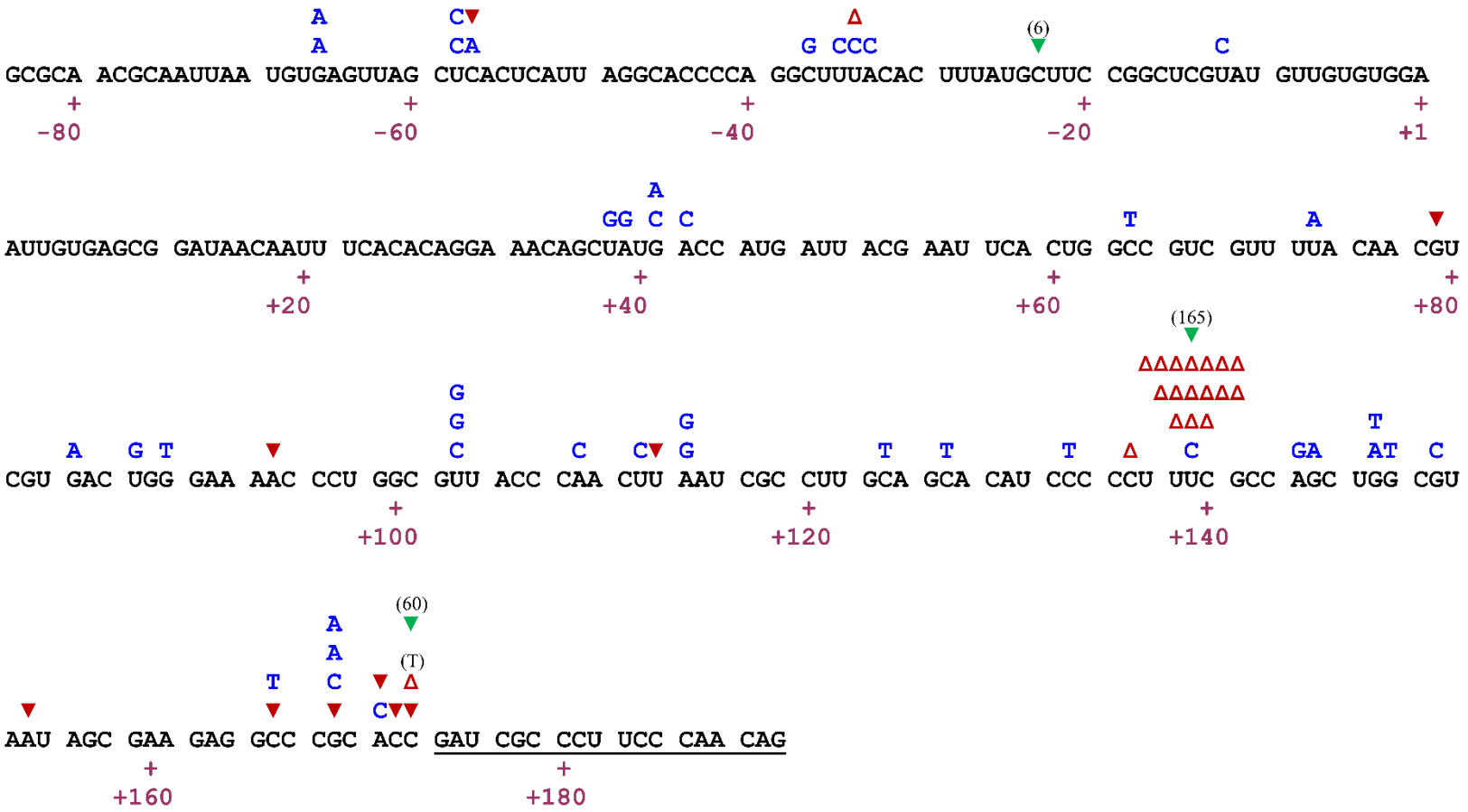


Figure 32. Spectrum of mutations induced by HIV-1_{BH10} RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the recombinant T7 RNA polymerase at pH 6.75 (Bis-Tris buffer) and in the presence of 1.5 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned, if not indicated). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

Table 17. Summary of error rates for RNA-dependent DNA synthesis catalyzed by HIV-1 RTs, and obtained with RNAs synthesized by the T7 RNAP under different assay conditions

Mutation type	HIV-1_{BH10} RT				$K65R/V75I_{ESP49}$ RT			
	Recombinant T7 RNAP (pH 7.9; 6 mM $MgCl_2$)		Recombinant T7 RNAP (pH 6.75; 1.5 mM $MgCl_2$)		Recombinant T7 RNAP (pH 6.75; 1.5 mM $MgCl_2$)		Commercial T7 RNAP [a] (pH 6.75; 1.5 mM $MgCl_2$)	
	No. of errors	Error rate	No. of errors	Error rate	No. of errors	Error rate	No. of errors	Error rate
All classes	**47**	**1/11872**	**74** [b]	**1/18204**	**59**	**1/29969**	**53**	**1/44074**
Base substitutions	**26**	**1/13199**	**42**	**1/19726**	**40**	**1/27186**	**32**	**1/44893**
Transitions	14 (54%)		25 (60%)		30 (75%)		19 (59%)	
Transversions	12 (46%)		17 (40%)		10 (25%)		13 (41%)	
Frameshifts	**21**	**1/23649**	**32**	**1/37467**	**19**	**1/82826**	**21**	**1/98999**
Insertions	10 (48%)		19 (59%)		16 (84%)		9 (43%)	
Deletions	11 (52%)		13 (41%)		3 (16%)		12 (57%)	
At runs [c]	10 (48%)	1/10602	21 (66%)	1/12188	15 (79%)	1/22397	11 (52%)	1/40348
At non-runs	11 (52%)	1/35509	11 (34%)	1/85726	4 (21%)	1/309435	10 (48%)	1/163514

[a] The T7 RNAP used in these assays was obtained from Promega, and buffers for the corresponding RNA synthesis reactions were prepared with PIPES. [b] Recombinant mutants are excluded from the analysis. [c] A run is considered when there is a row of three or more identical nucleotides.

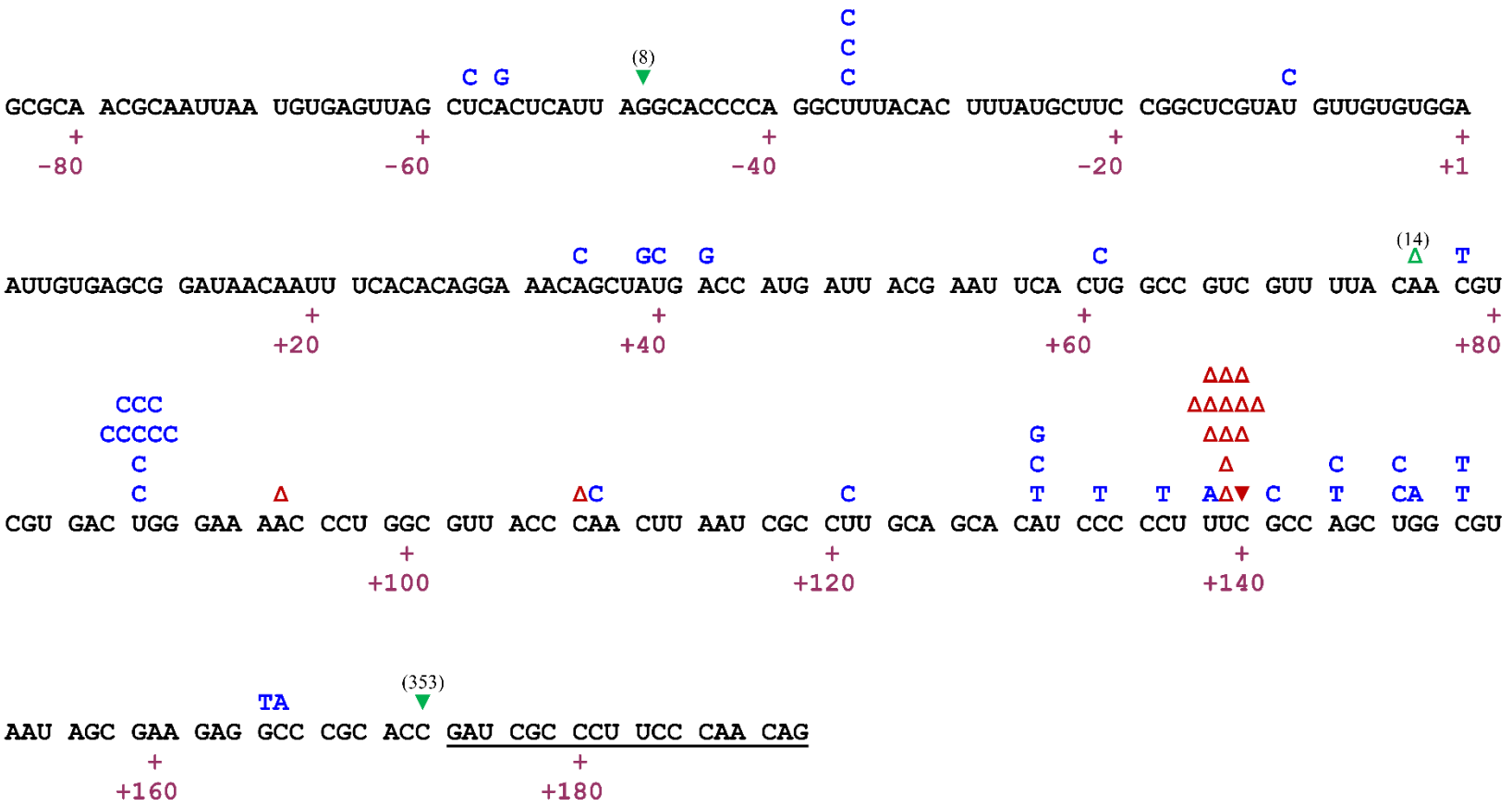


Figure 33. Spectrum of mutations induced by K65R/V75I$_{ESP49}$ RT in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the recombinant T7 RNA polymerase at pH 6.75 (Bis-Tris buffer) and in the presence of 1.5 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned), while open upright green triangles represent insertions of more than one nucleotide (the number of inserted bases is indicated in brackets). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used during cDNA synthesis.

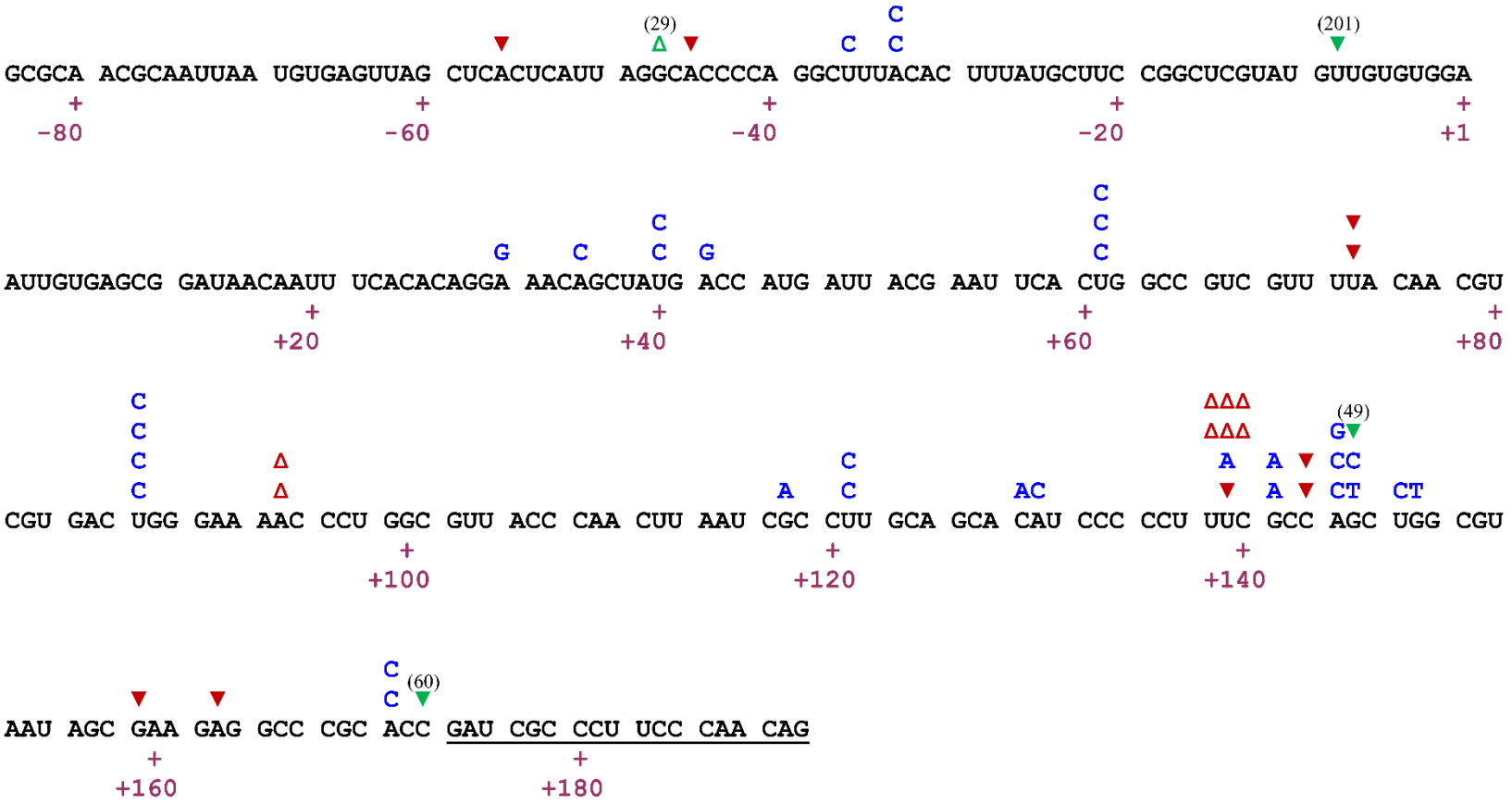


Figure 34. Spectrum of mutations induced by K65R/V75I RT_{ESP49} in M13mp2 *lacZα* forward mutation assays. The RNA used as template in the reverse transcription reaction was synthesized by the T7 RNA polymerase (Promega) at pH 6.75 (PIPES buffer) and in the presence of 1.5 mM $MgCl_2$. Single-nucleotide substitutions are indicated by the letter corresponding to the new base (in blue) above the template sequence of the *lacZα* target. Open upright red triangles represent insertions of one nucleotide (duplication of the base where the triangle is positioned), while open upright green triangles represent insertions of more than one nucleotide (the number of inserted bases is indicated in brackets). Inverted triangles indicate deletions of one (red) or more nucleotides (green). In the latter case, the triangle is positioned at the 3' end of the deletion and the number of deleted nucleotides is indicated above. Underlined bases represent the primer used for cDNA synthesis.

4.3. Determination of RNA-dependent DNA synthesis fidelity of RTs by deep sequencing experiments based on the use of Primer IDs

Although M13mp2-based forward mutation assays provide a broad assessment of RT's fidelity, allowing to compare data obtained from RNA- versus DNA-templated reactions, they are restricted to the detection of changes leading to a phenotypic effect. Thus, synonymous mutations and errors that do not affect the activity of the reporter gene (e.g., β-galactosidase activity in the case of *lacZα* assays) are not considered. In contrast, next-generation sequencing (NGS) technologies yield millions of read bases, providing depth in the analysis and allowing to detect all errors at any position.

In this Thesis, RNA-dependent DNA synthesis fidelity of RTs has been measured by using NGS platforms and protocols adapted for the use of Primer IDs, which are degenerated oligonucleotides used to initiate the synthesis of cDNA. This method provides a strategy to tag each cDNA molecule, and eliminates the bias generated by PCR amplification and sequencing errors, by creating a consensus between all cDNA molecules with the same barcode. Barcoded primers with twelve random nucleotides were used in reverse transcription reactions performed by HIV-1_{BH10} and K65R/V75I_{ESP49} RTs. Amplicon libraries were constructed from the cDNA samples, and sequenced by using MiSeq Illumina 300 system in paired-end (2×300). About 2×10^6 raw reads were obtained for each RT (**Table 18**), being the length of reads between 35 and 301 nucleotides.

Table 18. Raw reads and joined sequences by paired-end read merger (PEAR)

RTs	Original paired reads	Joined sequences	Joined sequences (%)
HIV-1_{BH10}	1,921,903	1,910,996	99.43
K65R/V75I_{ESP49}	1,743,634	1,733,270	99.40

Similar quality patterns of the samples were obtained for both enzymes (**Figure 35**). These patterns revealed a reduction in the quality of bases while advancing in the positions of the amplicon, following a general tendency in this kind of analysis. However, the lower quality of the distant bases is compensated by the joining of sequences read in paired-end during MiSeq sequencing. This process was performed with a fast and accurate paired-end read merger called PEAR.

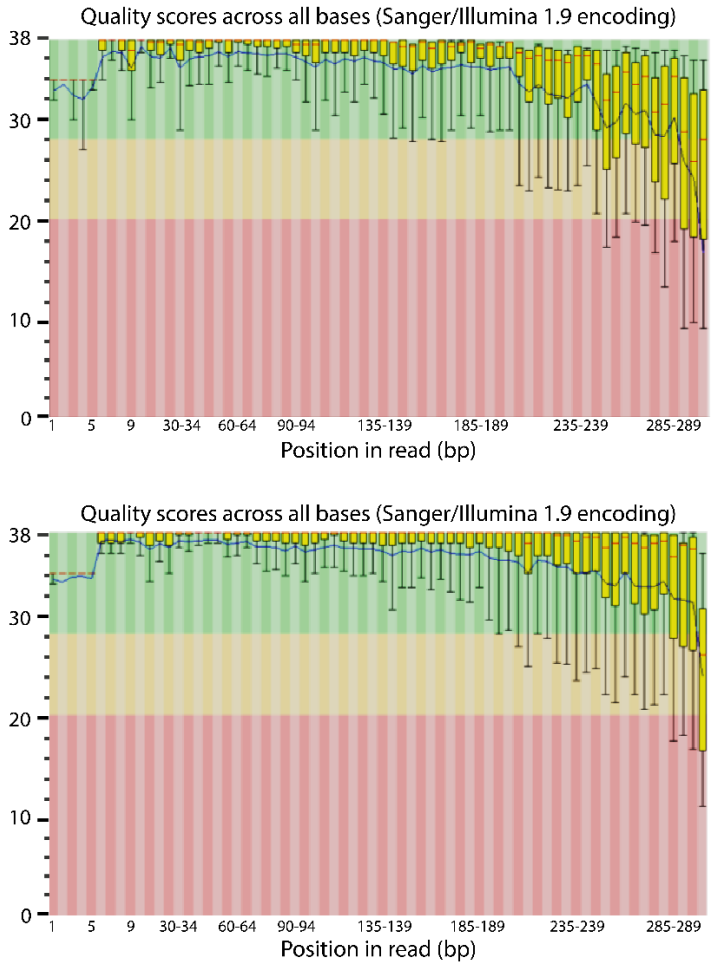


Figure 35. Representative quality per base boxplot of reads. HIV-1_{BH10} (above) and K65R/V75I_{ESP49} (below) RTs.

After joining paired-end sequences, >99% of the reads were recovered (**Table 18**). Consensus were determined from cDNA molecules with the same barcode, assuming the maximum limit of 10,000 reads and the lower limit established by the cutoff value (76 reads). This threshold yields few valid consensus (**Table 19**), since we obtained only 149 and 146 consensus for HIV-1_{BH10} and K65R/V75I_{ESP49} RTs, respectively. The high number of discarded consensus is mainly due to the high number of barcodes with very few reads. This biased distribution of reads per barcode may reflect the offspring barcodes generated by PCR and sequencing errors.

Table 19. Summary of obtained barcodes and the resulting consensus

RTs	Total barcodes	Maximum number of sequences per barcode	Consensus	Discarded
HIV-1_{BH10}	8,642	820,752	149	8,493
K65R/V75I_{ESP49}	8,289	530,475	146	8,143

Once the consensus sequences were generated, they were subjected to multiple alignment and compared to a reference sequence taken from LAV genome at the NCBI database. Variants found in the samples with respect to the reference sequence are indicated in **Table 20**, and show mutations that were introduced either by the eukaryotic RNAP II while synthesizing the RNA used as template, or by the RT during reverse transcription.

Table 20. Summary of mutations found in the consensus sequences for HIV-1_{BH10} and K65R/V75I_{ESP49} RTs

Enzyme	Position	Reference sequence	Mutated sequence	No. of errors	Frequency
HIV-1_{BH10} RT (146 consensus sequences)	61	AGGGGGGC	AGGGGGGGC	5	0.03425
	72	UAAAGG	TAAAAGG	2	0.01370
	114	UAG	TAAG	2	0.01370
	147	CAAAAAUG	CAAAAAATG	4	0.02740
	157	AGGGGGAAU	AGGGGGGAAT	2	0.01370
	241	AUUAG	ATTTAG	2	0.01370
	322	U	C	1	0.00685
K65R/V75I_{ESP49} RT (149 consensus sequences)	23	A	G	1	0.00671
	67	GCA	GCCA	2	0.01342
	122	AUG	ATTG	2	0.01342
	147	CAAAAAUG	CAAAAAATG	3	0.02013
	354	C	T	6	0.04027

Both mutational summaries showed the accumulation of one-adenine insertions at the homopolymeric sequence located around position +147, as well as the presence of C-to-T changes as unique base substitutions. Interestingly, all frameshifts observed for both RTs were one-nucleotide insertions. Different remarkable hotspots can be observed in the spectra of both enzymes. For example, one-nucleotide insertions of guanidine around position +61 in the case of HIV-1_{BH10} RT, or C-to-T transitions at position +354 in the case of K65R/V75I_{ESP49} RT.

Error rates were calculated for each type of error (**Table 21**), based on the number of consensus analyzed for each enzyme, and the number of positions were each mutation can be detected (355 sites for base substitutions and frameshifts, 83 for insertions at homopolymeric sequences, and 272 for insertions at heteropolymeric regions). NGS experiments revealed a 1.3-fold difference in fidelity between K65R/V75I_{ESP49} and HIV-1_{BH10} RTs in reactions carried out with RNA templates, in agreement with the results of M13mp2 forward mutation assays. These results confirm the lower accuracy of the HIV-1_{BH10} enzyme. HIV-1_{BH10} RT showed a higher overall error rate than the K65R/V75I_{ESP49} RT, mainly due to the high rate of frameshifts at homopolymeric sites. In contrast, the double mutant enzyme showed a higher tendency to introduce insertions at heteropolymeric

sites. In addition, K65R/V75I$_{ESP49}$ RT did not show a bias between the introduction of base substitutions and frameshifts.

Table 21. Summary of error rates for HIV-1$_{BH10}$ and K65R/V75I$_{ESP49}$ RTs for various classes of mutations, based on the use of Primer IDs in next-generation sequencing assays

Mutation type	HIV-1$_{BH10}$ RT		K65R/V75I$_{ESP49}$ RT	
	No. of errors	Error rate	No. of errors	Error rate
All classes	**18**	**1/2,879**	**14**	**1/3,778**
Base substitutions	**1**	**1/51,830**	**7**	**1/7,556**
Transitions	1 (100%)		7 (100%)	
Transversions	0 (0%)		0 (0%)	
Frameshifts	**17**	**1/3,049**	**7**	**1/7,556**
Insertions	17 (100%)		7 (100%)	
Deletions	0 (0%)		0 (0%)	
At runs[a]	13	1/932	3	1/4,122
At non-runs	4	1/9,928	4	1/10,132

[a] A run is considered when there is a row of three or more identical nucleotides.

Taken together, these results were in good agreement with those obtained from forward mutation assays (**Table 14**), supporting the validity and potential of this methodology. However, these error rates proceed from a low number of mutations, and it would be desirable to increase the number of consensus sequences to obtain a sufficiently robust analysis.

5. DISCUSSION

5.1. Fidelity of WT and drug-resistant HIV-2_{ROD} RTs

Lentiviral RTs (e.g., those of HIV-1, SIV or feline immunodeficiency virus) are known to be less accurate than oncoretroviral RTs, such as those of MLV or AMV (reviewed in Menéndez-Arias, 2009, 2013a). The study presented in this Thesis shows that WT HIV-2_{ROD} RT is only 1.6 times more faithful than the reference HIV-1_{BH10} RT. These differences in fidelity are similar to those reported for other primate lentiviral RTs when compared with HIV-1 group M subtype B polymerases, such as the RTs of strains BH10 or HXB2, using M13-based forward mutation assays. Thus, the HIV-1_{BH10} RT showed about two-fold decreased fidelity in comparison with a prototypic WT HIV-1 group O RT (Álvarez *et al*., 2009, 2013), while SIVmne and SIVagm RTs were only 1.3 and 1.8 times more accurate, respectively, than subtype B RTs (Stuke *et al*., 1997; Diamond *et al*., 2003).

Despite the relatively small differences in mutation rates, HIV-1 group M subtype B RTs emerge as the less faithful polymerases among those of primate lentiviruses while synthesizing DNA using DNA as template,. The higher intrinsic fidelity of HIV-2_{ROD} RT in comparison with HIV-1_{BH10} RT resulted from its lower nucleotide substitution error rate, a trend that was also observed with the SIVagm RT (Stuke *et al*., 1997). In contrast, our analysis showed that WT HIV-2_{ROD} RT has a relatively high frameshift error rate (5.7×10^{-5}), ranking among the highest reported for RTs of primate lentiviruses.

HIV-1 and HIV-2 RTs are important targets in the fight against AIDS. However, HIV-2 infected patients have fewer treatment options, since all antirretroviral drugs approved to treat HIV infections have been specifically designed to inhibit HIV-1 replication. Despite the inefficacy of NNRTIs, as well as the fusion inhibitor enfuvirtide and several protease inhibitors against HIV-2 (reviewed in Menéndez-Arias and Tözsér, 2008; Menéndez-Arias and Álvarez, 2014), both HIV-1 and HIV-2 strains display similar susceptibility to NRTIs. NRTIs constitute the backbone of current therapies against infections caused by all HIV isolates (Menéndez-Arias, 2013b). Challenging HIV with antiretroviral drugs facilitates the selection of strains containing mutations associated to drug resistance.

NRTI resistance is conferred by mutations that can act through two different mechanisms: (i) the exclusion pathway, that affects discrimination between canonical dNTPs and NRTIs, by reducing the incorporation efficiency of the latter; and (ii) the excision pathway, that promotes the removal of the already incorporated 3'-chain-terminating NRTI from the end of the blocked primer (**Figure 36**). Excision acts in a process similar to polymerization run in reverse and is mediated by a pyrophosphate donor, usually ATP. Unlike HIV-1, HIV-2 does not develop resistance to NRTIs

through the excision pathway, but relies exclusively on the nucleotide discrimination pathway towards the acquisition of resistance.

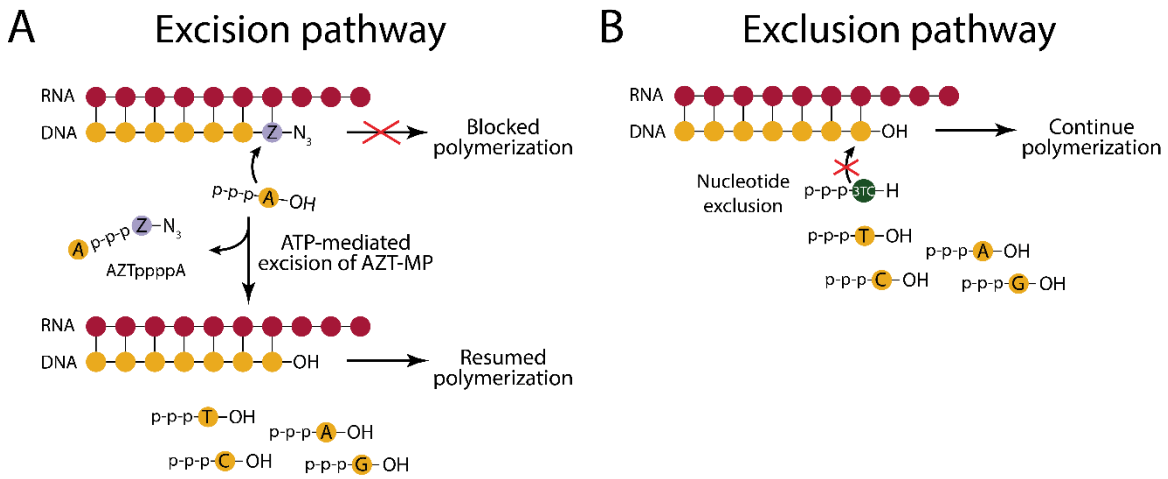


Figure 36. Mechanisms of NRTI resistance. (A) Excision pathway. AZT monophosphate (AZT-MP, purple) blocks polymerization once incorporated in the growing DNA chain, due to the absence of a 3'–OH end. The ATP-mediated removal of the already incorporated AZT-MP yields an AZTppppA excision byproduct and a 3'–OH end, facilitating the resumption of polymerization. **(B)** Exclusion pathway. Discrimination between natural dNTPs (yellow) and analogues, such as lamivudine (3TC, dark green) avoid certain drugs to enter in the growing DNA chain allowing the polymerization to continue. Both mechanisms yield a complex competent for polymerization. P, phosphate group; yellow circles with three phosphates, dNTPs. RNA is depicted with red circles, and DNA with yellow circles. Adapted from Delviks-Frankenberry *et al.* (2010).

In HIV-1, amino acid substitutions conferring resistance to NRTIs via the excision pathway are widely known as thymidine analogue resistance mutations (TAMs) (Larder, 1994). In patients receiving long-term therapy, TAMs appear in two clusters: TAM1 (M41L, L210W, and T215Y) and TAM2 (D67N, K70R, K219E/Q, and sometimes T215F) (Yahi *et al.*, 1999). In contrast, TAMs are rarely observed in cell culture after HIV-2 passage in the presence of AZT (Reid *et al.*, 2005), and are found with a very low prevalence in HIV-2 obtained from patients treated with AZT and other nucleoside analogues (Brandin *et al.*, 2003; Gottlieb *et al.*, 2009; Ntemgwa *et al.*, 2009).

In HIV-2, Met73 and to a lesser extent Ile75 play a critical role in the suppression of the excision pathway when TAMs are present in the RT (Álvarez *et al.*, 2018) (**Figure 37A**). The equivalent residues of the HIV-1 RT are Lys73 and Val75. All HIV-2 mutants having the substitution M73K in their RT were excision-proficient in biochemical assays. However, mutations selected under drug pressure frequently cause a loss of viral fitness. Thus, development of resistance through the excision pathway and involving the acquisition of M73K is highly unlikely, because HIV-2 variants carrying this substitution in the RT showed impaired viral replication (Álvarez *et al.*, 2018).

On the other hand, HIV-1 RT develops resistance to NRTIs via the exclusion pathway through the acquisition of amino acid changes found at the dNTP-binding site of the DNA polymerase

domain (e.g., K65R, L74V, V75I, and M184V). These changes antagonize the effects of TAMs while interfering with the RT's ability to discriminate between natural dNTPs and the triphosphate forms of NRTIs (reviewed in Menéndez-Arias, 2013b). In HIV-1 RT, K65R is selected in the presence of tenofovir, while Q151M was originally found in patients treated with AZT in combination with ddI or ddC. M184V appears under treatment with lamivudine (3TC) (reviewed in Menéndez-Arias, 2008). In the HIV-1 RT, the "Q151M complex" consists of a combination of five mutations (A62V, V75I, F77L, F116Y, and Q151M) (**Figure 37B**) that confers resistance to all approved NRTIs except tenofovir (Shirasaka *et al*., 1995; Hachiya *et al*., 2011). In contrast, in HIV-2 the "Q151M complex" is almost restricted to the unique presence of Q151M due to the different structure of the nucleotide binding site of the RT (reviewed in Boyer *et al*., 2006).

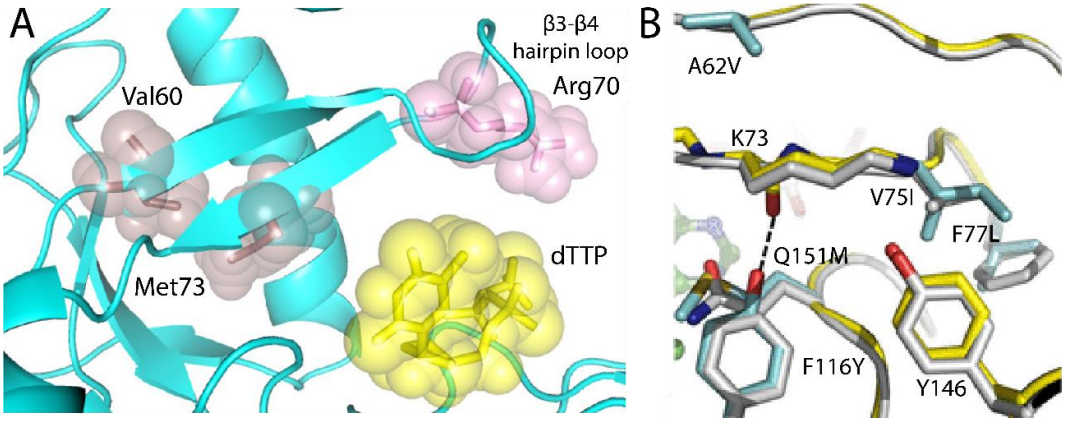


Figure 37. Structure of HIV-1 resistant RT. (A) Ribbon representation of the polypeptide backbone around the nucleotide-binding site of the mutant HIV-1 RT D67N/K70R/K73M. The side chains of Val60, Arg70, and Met73 (pink), and the incoming dNTP (yellow) are shown with stick and sphere representations. In this mutant RT, distance between the amido group of Arg70 and the hydroxyl group of the incoming dTTP reaches values up to 6.2 Å. This long distance is not observed in the double mutant D67N/K70R RT, where molecular dynamics simulations predict that those groups form a hydrogen bond. Adapted from Álvarez *et al*. (2018). **(B)** Structural superposition of WT HIV-1 RT (grey) and mutant HIV-1 RT (yellow) bearing the Q151M complex (A62V, V75I, F77L, F116Y, and Q151M) (cyan for mutated side chains). Adapted from Das *et al*. (2017).

The major mutations conferring resistance to NRTIs in HIV-2 are K65R, Q151M (sometimes accompanied by V111I), and M184V (Damond *et al*., 2005; Deuzing *et al*., 2015). Although K65R, Q151M and M184V might be selected under NRTI pressure in HIV-1 as well as in HIV-2, the triple mutant is characteristic of the HIV-2 RT since Q151M is the preferred mutation selected during treatment with AZT (Descamps *et al*., 2004; for reviews see Menéndez-Arias, 2013b; Menéndez-Arias and Álvarez, 2014). The combination of the three amino acid substitutions (K65R, Q151M, and M184V) confers classwide-nucleoside analogue resistance in HIV-2 RT (Smith *et al*., 2009). The study presented in this Thesis shows that effects on fidelity obtained with K65R (alone or in the context of a multidrug-resistant RT) were different in the HIV-2_{ROD} RT as compared with the HIV-1_{BH10} enzyme.

Interestingly, our studies revealed that there were relatively small differences in fidelity between the classwide NRTI-resistant K65R/Q151M/M184V$_{ROD}$ RT and the WT HIV-2$_{ROD}$ enzyme, despite the presence of K65R in the mutational complex. These effects were unexpected considering that K65R alone produced large increases in fidelity (>8-fold) in forward mutation assays carried out with HIV-1 RTs from phylogenetically distant strains such as HXB2 (group M subtype B) (Shah *et al*., 2000) or ESP49 (group O) (Barrioluengo *et al*., 2011). Furthermore, estimates of HIV-1 mutant frequencies obtained in cell culture after one round of replication were consistent with the enzymatic data and showed a 3.3-fold reduction for virus containing the K65R mutation in their RT, as compared with the virus carrying a WT RT (Mansky *et al*., 2003).

The results obtained with the triple mutant K65R/Q151M/M184V in the HIV-2$_{ROD}$ RT were in agreement with those obtained with the single mutant RT K65R$_{ROD}$. Thus, our results showed that K65R alone had almost no effect on the overall fidelity of the HIV-2$_{ROD}$ polymerase, although it produced a modest decrease in the base substitution error rates. Our data on the role of K65R in HIV-2 are in agreement with published misinsertion ratios obtained with WT HIV-2$_{ROD}$ RT and mutants K65R$_{ROD}$ and K65R/Q151M/M184V$_{ROD}$, performed in our laboratory using gel-based assays (Álvarez *et al*., 2017). Differences in misinsertion ratios between WT and mutant RTs bearing K65R were small and non-significant, except for the case of A opposite A, which was about 4 times less efficient for the single mutant K65R than for the WT HIV-2$_{ROD}$ RT. However, dATP incorporation was very inefficient in the sequence context used in the study, suggesting that this type of errors would be rare during reverse transcription.

Moreover, classwide NRTI-resistant K65R/Q151M/M184V$_{ROD}$ RT and the WT HIV-2$_{ROD}$ RT did not show significant differences in their mispair extension ratios, when the kinetic parameters could be reliably determined (Álvarez *et al*., 2017). The K65R$_{ROD}$ mutant RT showed increased fidelity in comparison with the WT HIV-2$_{ROD}$ RT only in extension reactions with template-primers having a G:G mismatch at their 3' ends. However, in this case the catalytic efficiency of nucleotide incorporation was very low (in the order of 10^{-6} μM^{-1} s^{-1}).

On the other hand, the effects of K65R on the evolutionary rate of HIV-2 have not been assessed in previous studies. However, a recent work has shown the quick reversion of the mutation when present in SIVmac239 infecting pigtailed macaques (Lloyd *et al*., 2016). Interestingly, WT and mutant SIVmac239 showed similar mutation rates in cell-based assays although K65R-containing viruses showed slightly reduced variability. The differences in mutation rates were estimated to be less than 1.3-fold, but considered statistically significant (Lloyd *et al*., 2016). It is noteworthy to mention that mutant MLV carrying the equivalent substitution in its RT (i.e., K103R) was found to be less infectious than the WT. However, no differences were found in assays measuring the frequency of *lac*Z inactivation after one round of replication in cells infected with WT and mutant

viruses (Halvas *et al*., 2000), thereby suggesting that K103R had no effect on the intrinsic fidelity of the MLV RT.

Although several studies carried out with HIV-1 RT variants have shown that K65R and other mutations conferring resistance to nucleoside analogues (e.g., M184V) increase their DNA polymerase fidelity, there are examples of NRTI resistance mutations having little effect on the accuracy of HIV-1 RT (e.g., L74V, Q151M) (reviewed in Menéndez-Arias, 2009, 2013a). Our results show that in HIV-2, classwide NRTI resistance may not be associated with significant changes in the intrinsic fidelity of the RT, although a lower fitness is expected for drug-resistant viruses as a result of the lower catalytic efficiency of their RT. Therefore, the relationship between drug resistance and increased fidelity cannot be established, particularly for the major mutations found in HIV-2 strains resistant to NRTIs.

The increased intrinsic fidelity conferred by K65R appears to be a specific characteristic of HIV-1 RT and not shared by RTs of the HIV-2/SIV lineage. A structural justification of these differences is very difficult due to the lack of structural information for HIV-2 RT. However, a number of amino acid sequence differences found in the β3-β4 hairpin loops (residues 63–75) of HIV-1 and HIV-2 RTs (**Figure 38**) could have a significant impact on its conformation and mobility, thereby affecting neighbouring contacts between the side-chains of Lys or Arg65. In support of this proposal, crystal structures of ternary complexes composed of mutant K65R HIV-1_{BH10} RT, dsDNA, and incoming nucleotides (i.e., dATP or tenofovir-diphosphate) have shown stacking interactions between guanidinium groups of Arg65 and Arg72 that restrict the conformational adaptability of both residues (Das *et al*., 2009).

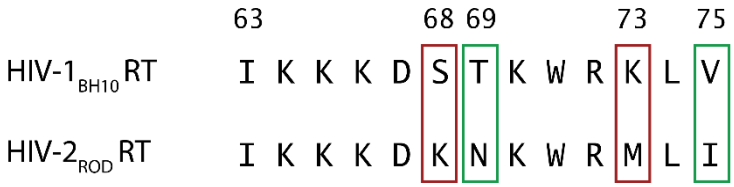


Figure 38. Alignment of β3-β4 hairpin loop (residues 63-75) amino acid sequences of HIV-1_{BH10} and HIV-2_{ROD}-SIV RTs. Conservative and non-conservative changes are indicated in green and red boxes, respectively.

The amino acid sequences of the β3-β4 hairpin loops of HIV-1 group M subtype B and group O RTs are identical (IKKKDSTKWRKLV), but different from those found in HIV-2_{ROD} and SIVmac239 RTs (IKKKDKNKWRMLI) (**Figure 38**). The distribution of charged residues along the hairpin loop sequence is rather different in both pairs of RTs and it is tempting to speculate that substitutions of basic residues in these structures could have an impact on fidelity. Nevertheless, how interactions between hairpin loop residues may influence the intrinsic fidelity of the enzyme is a question open for future research.

5.2. RNA-dependent DNA synthesis fidelity of retroviral RTs

5.2.1. Fidelity of retroviral RTs in reactions carried out with RNA templates

Previous studies on the accuracy of different retroviral RTs, such as those of WT HIV-1_{BH10}, HIV-1_{ESP49}, MLV and AMV, as well as mutant RTs $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$, while synthesizing DNA using DNA templates revealed that there were large differences in fidelity between all of them (reviewed in Menéndez-Arias, 2009, 2013a). In contrast, the results obtained for these RTs in the present study show that the intrinsic fidelity of RNA-dependent DNA synthesis varies within the small range of 2.5×10^{-5} to 3.5×10^{-5}, in assays carried out with RNA transcribed by a commercial T7 RNAP at pH 7.9 and in the presence of 6 mM Mg^{2+}. The comparison of mutant frequencies obtained with M13mp2 *lac*Zα forward mutation assays illustrates the large variation in fidelity observed between different RTs depending on the template used in the reaction (DNA or RNA) (**Figure 39**).

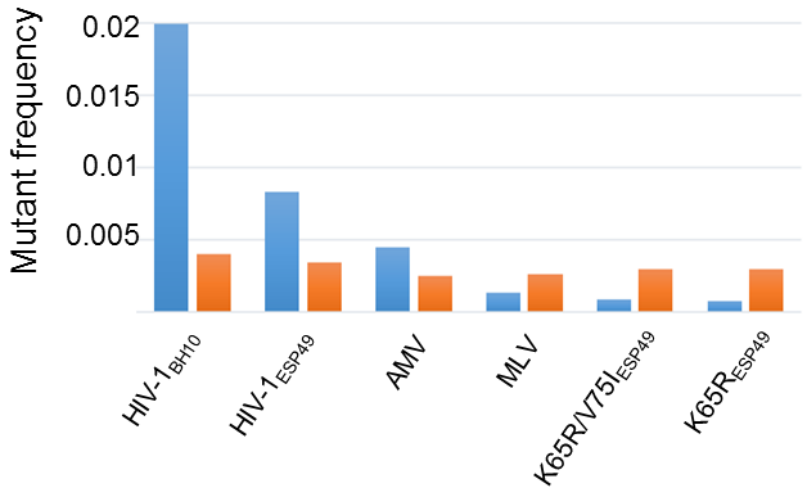


Figure 39. Comparison of RNA- and DNA-dependent DNA synthesis fidelities of retroviral RTs. Estimates of fidelity of RNA-dependent DNA synthesis are based on mutant frequencies obtained with the adapted M13mp2 *lacZα* forward mutation assay, reported in this Thesis (orange bars). Blue bars represent previously reported values, obtained using M13mp2-based assays measuring the accuracy of DNA-dependent DNA synthesis of HIV-1_{BH10} (Álvarez *et al*., 2017), HIV-1_{ESP49} (Álvarez *et al*., 2009), AMV (Roberts *et al*., 1988), and MLV, as well as mutant $K65R_{ESP49}$ and $K65R/V75I_{ESP49}$ RTs (Barrioluengo *et al*., 2011).

While HIV-1_{BH10} RT is >15-fold less accurate than mutant HIV-1_{ESP49} RTs K65R and K65R/V75I in DNA-dependent DNA synthesis reactions (Barrioluengo *et al*., 2011), differences between the most and least faithful RTs in assays carried out with RNA templates fall within less than two-fold (**Table 14**). Nonetheless, HIV-1_{BH10} RT remains as the least faithful enzyme in both types of assays.

All results reported in **Figure 39** were obtained in our laboratory under the same experimental conditions, except in the case of mutation frequencies determined for AMV RT with DNA-dependent DNA synthesis fidelity assays (Roberts *et al*., 1988). Our previous studies measuring the fidelity of DNA synthesis on DNA templates had shown that for the same RT, the variability in mutant frequencies obtained in forward mutation assays was under 30% for HIV-1_{BH10} and HIV-1_{ESP49} RTs (Álvarez *et al*., 2009, 2013, 2017).

The narrow window of variation in mutant frequencies or overall error rates (around two-fold) observed for retroviral RTs on RNA templates was also found when changing transcription conditions. Thus, in assays performed with RNA templates synthesized by a recombinant T7 RNAP using reaction conditions that favoured the accuracy of the transcription (i.e., pH 6.75/1.5 mM Mg^{2+}), overall error rates were in the range of 3.3×10^{-5} to 5.5×10^{-5}. The lower value of error rate obtained on RNA templates was 2.3×10^{-5}, and it was obtained with the K65R/V75I_{ESP49} RT using RNA templates synthesized by a commercial T7 RNAP at pH 6.75/1.5 mM Mg^{2+}. In contrast, intrinsic error rates of RTs in DNA-dependent DNA synthesis can be as low as 6.3×10^{-6} (e.g., for K65R_{ESP49} and K65R/V75I_{ESP49} RTs) (Barrioluengo *et al*., 2011). Thus, in RNA-templated assays error rates were more than three-fold higher than the lowest values obtained with DNA templates.

Structural studies with HIV-1 RT bound to RNA/DNA template-primers (Sarafianos *et al*., 2001; Lapkouski *et al*., 2013; Das *et al*., 2014; Tian *et al*., 2018) have shown that the RT adopts a similar conformation than in complexes containing DNA/DNA hybrids. However, in the structures containing RNA/DNA complexes, there are additional contacts involving residues of the p51 subunit and 2' −OH groups of the RNA template. These interactions contribute to the higher affinity of HIV-1 RT for RNA/DNA hybrids obtained in biochemical studies (Bohlayer and DeStefano, 2006). Differences in the interaction between RTs and RNA/DNA complexes *versus* DNA/DNA template-primers could be potentially responsible for the different types of errors obtained in assays measuring fidelity of RNA- *versus* DNA-dependent DNA synthesis.

5.2.2. Mechanistic insights of RT-catalyzed DNA polymerization reaction with RNA or DNA templates

The mutational specificity of the studied RTs was determined after sequencing the *lacZα* gene of mutants generated in M13-based assays. By analysing the mutational spectra induced by the different RTs we also obtained relevant information on preferred hotspots and types of errors on different templates (i.e., RNA *versus* DNA), and the contribution of specific errors likely made by T7 RNAP and found in RNA templates used in the assays.


GCGCA ACGCAAUUAA UGUGAGUUAG CUCACUCAUU AGGCACCCCA GGCUUUACAC UUUAUGCUUC CGGCUCGUAU GUUGUGUGGA
-80 -60 -40 -20 +1
AUUGUGAGCG GAUAACAAUU UCACACAGGA AACAGCUAUG ACC AUG AUU ACG AAU UCA CUG GCC GUC GUU UUA CAA CGU
+20 +40 +60 +80
CGU GAC UGG GAA AAC CCU GGC GUU ACC CAA CUU AAU CGC CUU GCA GCA CAU CCC CCU UUC GCC AGC UGG CGU
+100 +120 +140
AAU AGC GAA GAG GCC CGC ACC GAU CGC CCU UCC CAA CAG
+160 +180
GCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA
-80 -60 -40 -20 +1
ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCTATG ACC ATG ATT ACG AAT TCA CTG GCC GTC GTT TTA CAA CGT
+20 +40 +60 +80
CGT GAC TGG GAA AAC CCT GGC GTT ACC CAA CTT AAT CGC CTT GCA GCA CAT CCC CCT TTC GCC AGC TGG CGT
+100 +120 +140
AAT AGC GAA GAG GCC CGC ACC GAT CGC CCT TCC CAA CAG CTG CGC
+160 +180 +197

Figure 40. Comparison of mutational spectra induced by retroviral RTs during RNA- and DNA-dependent DNA synthesis. (Up) Combined mutational spectra derived from RNA-dependent DNA synthesis reactions including all RTs analyzed in this Thesis (**Figures 22 to 27**). RTs included are those of HIV-1_{BH10} (pink), HIV-1_{ESP49} (purple), AMV (red), and MLV (green), and mutants K65R/V75I_{ESP49} (light blue) and K65R_{ESP49} (brown). **(Bottom)** Combined mutational spectra induced during DNA-dependent DNA synthesis reactions, and taken from previously published reports (Roberts *et al*., 1989; Álvarez *et al*., 2009; Barrioluengo *et al*., 2011). RT color codes are the same as above. For all RTs and mutational spectra, single-nucleotide substitutions are indicated by the letter corresponding to the new base above the template sequence of the *lacZα* target. Open upright triangles represent insertions and inverted triangles indicate deletions. Indels are represented by triangles which are positioned at the 3' end of the frameshift with the number of inserted/deleted nucleotides indicated between parentheses. One-nucleotide insertions correspond to the duplication of the base where the triangle is positioned, if not specified.

The mutational spectra of the six RTs studied (collected in **Figure 40**) show that RTs are prone to introduce errors when copying RNA or DNA templates at positions −36/−34 (mostly, U-to-C or T-to-C transitions), +139/+149, and within nucleotides +165 and +171. In the latter case, transversions are largely predominant with the DNA template, while transitions, frameshift errors and large deletions dominated the RNA-dependent mutational spectra. In addition, we observed a different distribution of hotspots when comparing spectra obtained with RNA or DNA templates. The DNA-dependent mutational spectra showed scattered hotpots at positions −66, +108, +118, +136, and clustered at nucleotides +87 to +92. In contrast, the RNA-dependent spectra contained hotspots at positions −7, +73, +87, +109, and +121. Among them, the hotspot at position +73 was represented mainly by U-to-A transversions and one-nucleotide frameshifts.

The mutational spectra induced by RTs during RNA-dependent DNA synthesis contain a prominent hotspot at the homopolymeric region located at +137/+139. Errors at this position consist of insertions of one T in a run of three thymidines. These errors were previously detected by Boyer *et al*. (1992) in spectra obtained with HIV-1 and AMV RTs, and attributed to errors made by the T7 RNAP. Therefore, those errors would be already found in the RNA template used by retroviral RTs in the cDNA synthesis reaction. Our results are consistent with that proposal, since we also found one-nucleotide insertions at this site in all of the analysed mutational spectra, without statistically significant differences between them (**Table 13**). Interestingly, the insertions were not detected at positions +137/+139 in the mutational spectrum induced by HIV-1 RT using *lacZα* RNA templates synthesized with the T3 RNAP (Ji and Loeb, 1992), supporting the notion that this particular error is specifically made by the bacteriophage T7 RNAP. In general, it appears that monotonous runs of adenines in DNA are universal hotspots for RNAP slippage events during transcription. Adenine runs have been found to be slippery *in vitro*, in model systems in *Escherichia coli* and yeast, and in genome-wide transcriptome analysis of epimutations in *Caenorhabditis elegans* (reviewed in Gordon *et al*., 2015).

Another important hotspot represented by many large deletions occurs at positions +167/+173, at the 3' end of the primer used for cDNA synthesis by the retroviral RTs. Deletions were more frequent at positions that correspond to the incorporation of the first and second nucleotides (i.e., +172 and +173), and probably result from either inefficient extension of the primer during DNA synthesis or aberrant RT/template-primer interactions. Remarkably, frameshift errors near the 3' end of the primer used in cDNA synthesis had been previously reported by Boyer *et al.* (1992), although in that study the hotspot induced by HIV-1 RT appeared at position +158. Interestingly, the authors also noted that this hotspot was absent from the mutational spectra generated with AMV RT. In our experiments, the large deletions at the 3' end of the primer were found only in the mutational spectra induced by HIV-1_{BH10}, HIV-1_{ESP49} and the K65R_{ESP49} RT, but not with oncoretroviral RTs such as MLV or AMV RTs and the HIV-1_{ESP49} mutant K65R/V75I. The presence of this particular hotspot could be partly explained by the relatively high template-primer dissociation rate of HIV-1 RT during the first nucleotide incorporation events (Majumdar *et al*., 1988).

5.2.3. Influence of transcription inaccuracy in forward mutation assays

Our results show that M13mp2 assays yield values of overall error rates for six retroviral RTs in the range of 10^{-5}, measured in the presence of RNA templates synthesized by the T7 RNAP. Differences in fidelity between the enzymes (i.e., WT HIV-1_{BH10}, HIV-1_{ESP49}, MLV and AMV, and mutant RTs K65R_{ESP49} and K65R/V75I_{ESP49}) were around two-fold. This similarity in accuracy observed between different RTs while reverse transcribing RNA molecules is likely due to transcription errors affecting the quality of the RNA template used in the reactions. Transcription errors or epimutations (i.e., RNA mutations resulting from inaccuracy during transcription) are very difficult to quantify and large differences in error rates have been reported by several groups. The absence of a standardized method hampers the comparison of data coming from experiments that vary enormously depending on the studied RNAP (e.g., single- or multi-subunit, from unicellular or pluricellular, prokaryotic or eukaryotic organisms) and on the used technical approach (i.e., *in vitro* or *in vivo*, with methodologies that differ in drawbacks and advantages) (**Table 22**).

Even with these disparities, studies with multiple RNAPs rendered rough estimates of transcriptional mutagenesis in the order of 10^{-5} transcription errors per nucleotide (Remington *et al*., 1998; Huang *et al*., 2000; Imashimizu *et al*., 2013; Traverse and Ochman, 2016; Reid-Bayliss and Loeb, 2017; reviewed in Ninio, 1991; Magnuson *et al*., 2016), in agreement with the threshold observed for the synthesis of DNA on RNA templates with different RTs.

Gel-based assays showed that T7 RNAP exhibited an average error frequency of 5×10^{-5} obtained from measurements of nucleotide misincorporation ratios (Huang *et al*., 2000), in good

agreement with the average base substitution error rate of 3×10^{-5} determined by Remington *et al*. (1998) with a codon reversion assay. In contrast, Sultana *et al*. (2017) obtained a transcription error rate of 2×10^{-6}, determined as an average of 11 of the 12 possible nucleotide misinsertion efficiencies (f_{ins}) at a given position (i.e., 4 nucleotides can occupy a specific template position, each one yielding 1 correct and 3 incorrect incorporations).

However, in both reports the comparison between incorrect and correct nucleotides was limited to a relatively small number of sequences. In addition, Huang *et al*. (2000) compared all the misincorporation rates, by using a normal transcript extension rate for the T7 RNAP of 60 bases/s, taken from Ikeda and Richardson (1987). On the other hand, Sultana *et al*. (2017) compared the 11 catalytic efficiencies for the incorporation of incorrect nucleotides taking as a reference the catalytic efficiency (k_{pol}/K_d) of the incorporation of U opposite a template dA (i.e., $2 \pm 0.5\times10^{-6}$), instead of comparing the catalytic efficiency of each incorrect nucleotide with that of the corresponding correct one.

Table 22. Summary of transcription and reverse transcription error rates

Technique	Error rate [a]	Enzyme analyzed [b]	Reference
Screening by electrophoretic mobility			
SSCP	$2.5-5.6\times10^{-5}$ (O)	RNAP II + RT [RNA]	O'Neil *et al*., 2002
	$2.4-6.7\times10^{-5}$ (O)	RT [DNA]	O'Neil *et al*., 2002
Kinetics analysis			
Nucleotide incorporation	2×10^{-6} (BS)	T7 RNAP	Sultana *et al*., 2017
Deep sequencing methods			
Rep-Seq	2.2×10^{-6} (BS)	*C. elegans* transcripts	Gout *et al*., 2013
	1.2×10^{-6} (F)	*C. elegans* transcripts	Gout *et al*., 2013
Primer ID	2.7×10^{-4} (O)	T7 RNAP + HIV-1 RT [RNA]	Yasukawa *et al*., 2017
	1.3×10^{-4} (O)	RNAP II + MLV RT [RNA]	Zhou *et al*., 2015
	1.1×10^{-4} (BS)		
HiRes-Seq	$10^{-4}-10^{-5}$ (BS)	*E. coli* RNAP *in vitro* and *in vivo*	Imashimizu *et al*., 2013
Cir-Seq	$2.4-3.3\times10^{-4}$ (BS)	Poliovirus population	Acevedo *et al*., 2014
	$3.3-8.2\times10^{-5}$ (BS)	RNAPs of *E. coli*, *B. aphidicola*, and *C. ruddii*	Traverse and Ochman, 2016
	$3.9\times10^{-6}-1.7\times10^{-5}$ (BS)	*S. cerevisiae* transcripts	Gout *et al*., 2017
ARC-Seq	2.0×10^{-5}	T7 RNAP *in vitro*	Reid-Bayliss *et al*., 2017
SMRT-Seq	$5.6-7.5\times10^{-5}$ (O)	AMV and MLV RT [RNA] + T7 RNAP	Potapov *et al*., 2018
	$5.2-8.4\times10^{-5}$ (O)	AMV and MLV RT [DNA]	

SSCP, single-stranded conformation polymorphism; Rep-Seq, replicated sequencing; HiRes-Seq, high-resolution RNA sequencing; Cir-Seq, circular resequencing; ARC-Seq, accurate RNA consensus sequencing; and SMRT-Seq, single molecule real-time sequencing.

[a] Error rates for base substitutions (BS), frameshifts (F) or overall (O) error rate. [b] In brackets, DNA and RNA indicates the template used in the reaction, related to the corresponding error rate.

Based on deep sequencing technologies, Traverse and Ochman (2016) found conserved transcription error rates (ranging from 3.3×10^{-5} to 8.2×10^{-5} per nucleotide) across different species of bacteria. These error rates were determined by circular resequencing (Cir-Seq) under different growth conditions (minimal and complex medium), and by considering either messenger or ribosomal RNA sequences. A scheme of the Cir-Seq method is depicted in **Figure 41**, and compared with other deep-sequencing protocols applied to the determination of transcription error rates.

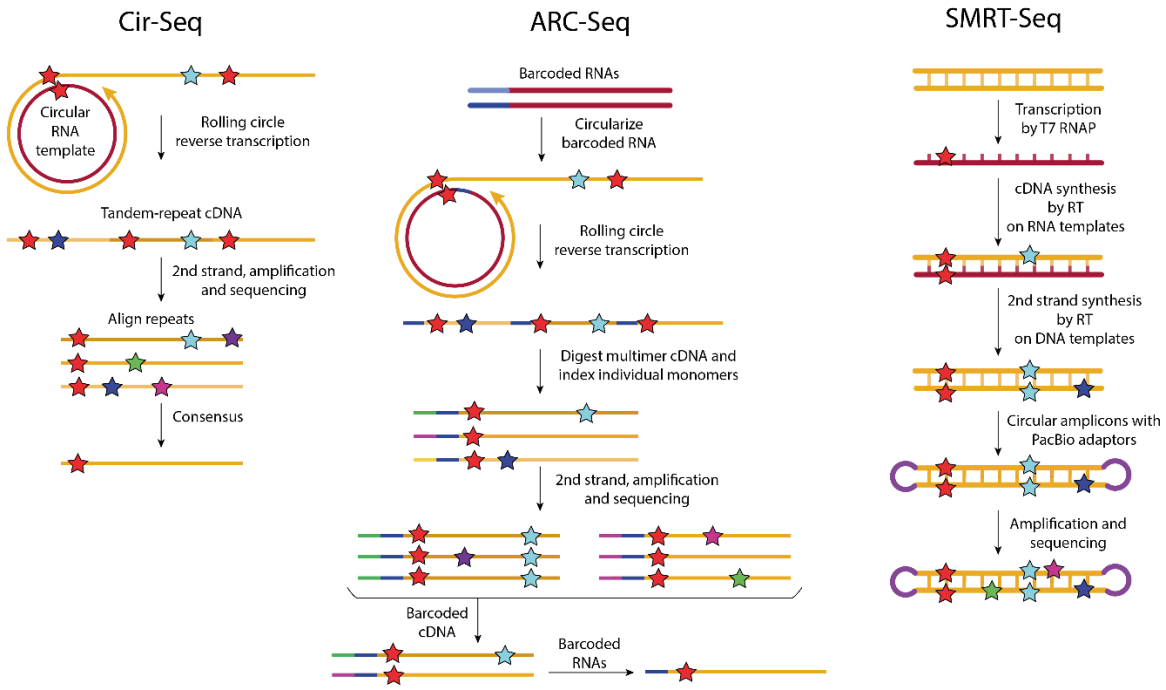


Figure 41. Summary of techniques used for the study of transcription and reverse transcription error rates. (A) Circular resequencing (Cir-Seq) (Acevedo and Andino, 2014). Rolling-circle reverse transcription of circularized RNA templates yields tandem-repeated copies of cDNA, which are cloned to generate a library of dsDNA molecules containing sequencing platform adapters. The tandemly linked reads are aligned, and consensus are obtained after computational processing, excluding sequencing and RT errors. **(B)** Accurate RNA consensus sequencing (ARC-Seq) (Reid-Bayliss and Loeb, 2017). Barcoded RNAs are circularized and subjected to rolling-circle reverse transcription, generating a multimeric cDNA from each RNA molecule. Then, they are restricted into monomers and tagged with a unique index, amplified and sequenced. Consensus of cDNA molecules with the same barcode, and then of cDNA consensus with the same RNA barcode allows to genuinely identify transcription and reverse transcription errors. **(C)** Single Molecule Real-Time sequencing (SMRT-Seq) (Potapov *et al*., 2018). DNA templates are transcribed by T7 RNAP. The obtained RNA is used as template of reverse transcription by RT that also synthesizes the 2nd strand. Adaptors circularize the amplicons before amplification. Transcription and reverse transcription errors are observed in both strands, while errors made by RT during 2nd strand synthesis are uniquely found in one strand producing a mismatch, as amplification and sequencing errors. Only errors found in both strands of DNA are counted. Errors of transcription (red), reverse transcription (light blue), 2nd strand synthesis (dark blue), and amplification and sequencing (green, and pink) are shown as stars.

Recently, Reid-Bayliss and Loeb (2017) developed an outbreaking method called accurate RNA consensus sequencing (ARC-seq) that allowed to distinguish between transcription errors and

those made by the RT during first-strand (cDNA) synthesis (**Figure 41**). This technique is based on a new protocol for library preparation, and allowed the authors to set the value of transcription mutagenesis at 2×10^{-5} for T7 RNAP. This value was consistent with the error rates obtained in our study, suggesting that T7 RNAP might be responsible for the inaccuracy threshold observed in our assays. Despite the scope of the technique and the broad specific information that could be obtained with this approach, authors restricted the study to transcription mutagenesis, and reverse transcription error rates were not provided.

According to these studies, we reasoned that T7 RNAP was imposing a level of transcriptional mutagenesis that for all the RTs was reflected in a threshold of around 10^{-5} errors per nucleotide (**Tables 14 and 17**). This inaccuracy in the synthesis of the RNA used as template hampered the evaluation of the fidelity of RTs during reverse transcription. We modified the conditions of the transcription reaction in order to alter the fidelity of T7 RNAP. Our results showed that indeed the accuracy of the T7 RNAP can be manipulated to some extent by changing transcription conditions, although reductions of the inaccuracy threshold were modest.

T7 RNAP synthesizes more faithful RNA at lower pH (6.75 *vs.* 7.9) and Mg^{2+} concentration (1.5 mM *vs*. 6 mM). This increase in fidelity using the lower pH conditions has also been previously reported for several DNA polymerases (e.g., Taq DNA polymerase, exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I, human DNA polymerase α and HIV-1 RT) (Eckert and Kunkel, 1990, 1993a, 1993b). Thus, increases of fidelity of up to 50-fold have been reported at pH 6.2 compared to pH 9.8 for base substitutions using the Klenow fragment of *E. coli* DNA polymerase I (Eckert and Kunkel, 1993a). Authors attributed these effects to altered template binding properties of the enzyme at a lower pH. In the case of T7 RNAP, nucleotide incorporation studies with nonpolar thymidine analogues have shown that hydrogen bonds are important for polymerization efficiency due to their role in stabilizing the closed ternary complex of the polymerase, but electrostatic interactions with the minor groove of the substrate were critical for fidelity (Ulrich and Kool, 2011).

On the other hand, it has been shown that HIV-1 RT and Taq DNA polymerase display higher fidelity at low concentrations of Mg^{2+} (e.g., 0.25 mM), although this property was controversial in experiments carried out with the MLV RT (Eckert and Kunkel, 1990; Achuthan *et al*., 2014; Okano *et al*., 2018). Despite improvements in fidelity obtained by reducing pH and Mg^{2+} concentrations, there are obvious limitations to this approach due to the loss of catalytic activity of the enzymes when departing from the optimal conditions for their polymerization activity.

High-fidelity transcription conditions (pH 6.75 and 1.5 mM Mg^{2+}) were used to transcribe the RNA that serves as template in M13mp2 forward mutation assays, achieving a slight reduction in the error rates of HIV-1_{BH10} and K65R/V75I_{ESP49} RTs. Although modest, the changes in fidelity observed for the same RT under different conditions of transcription could be attributed to the

accuracy of the bacteriophage T7 RNAP. For example, the error rate calculated for HIV-1_{BH10} RT increased 2.4-fold (from 3.5×10^{-5} to 8.4×10^{-5}) only by changing the commercial T7 RNAP for the recombinant enzyme.

In this Thesis, we also measured the RNA-dependent DNA synthesis fidelity based on the use of Primer IDs (i.e., degenerate oligonucleotides) for the synthesis of cDNA. This technique allows to discard amplification and sequencing errors, and provides a broad assessment of fidelity determined from multiple template sequences. Overall error rates of 3.5×10^{-4} and 2.6×10^{-4} mutations per nucleotide were obtained for HIV-1_{BH10} and K65R/V75I_{ESP49} RTs. This result was in agreement with previous data that focused on the study of RTs accuracy using the HIV-1 protease coding region as a target (Zhou *et al*., 2015), which rendered an overall error rate of 1.3×10^{-4} for a commercial RT from MLV.

In our experiments, as well as in the study of Zhou *et al*. (2015), HIV-1 (LAV) RNA genome was used as template. These templates are expected to be more accurate than those made by T7 RNAP, because they have been extracted from 8E5 cells, which constitute a cell line derived from human T cells. 8E5 cells contain only one integrated copy of an HIV-1 provirus, whose transcription is controlled by a eukaryotic RNAP (i.e., RNAP II). It has been reported that prokaryotic and eukaryotic RNAPs show several proofreading mechanisms, associated to intrinsic capacities and extrinsic assistant factors (Knippa and Peterson, 2013; Bubunenko *et al*., 2017; Mellenius and Ehrenberg, 2017; Mishanina *et al*., 2017; reviewed in Gamba and Zenkin, 2018). However, the high number of homopolymeric sites found in the target region of the protease used in our experiments (**Figure 17**) is expected to contribute to increase its overall error rate. In addition, the low number of consensus and mutations obtained (**Table 20**) limit the significance of these data. A more extensive analysis with a greater number of consensus sequences is necessary to obtain robust data on transcription and reverse transcription accuracy *in vitro*.

The transcriptional threshold can be modestly reduced, but it is clear that it has a measurable influence on the assessment of genuine errors made by RTs while copying RNA. Nevertheless, very often, studies have assumed that transcriptional mutagenesis has a relatively small effect in measuring the RNA-dependent DNA polymerase fidelity of RTs, suggesting that errors made by RNAPs are negligible, and do not significantly affect estimates of error rates induced by RTs. This widespread idea does not rely on conclusive results and is mainly based on studies that assume the equality of RNA- and DNA-dependent fidelities of RTs, or derived from misinterpretations of data provided by several studies, for example those of O'Neil *et al.* (2002).

The mutation rate for HIV-1 replication, which represents the combined error rate for RT and the human RNAP II, is approximately 2×10^{-5} per nucleotide per replication cycle, a rate that is similar to other retroviruses (Hu and Hughes, 2012; and references therein). The relative

contribution between the human RNAP II and the RT in the diversity of HIV-1 was determined by analysing the mutations produced in the HIV-1 long terminal repeats (LTRs) (O'Neil *et al*., 2002). Similar number of mutations were found in both LTRs (caused either by RNAP II during transcription or by RT during minus strand synthesis) and in one LTR (owing to the activity of RT while synthesizing the plus strand). Authors reported error rates of mutations found at both LTRs (made by both enzymes) of 2.5×10^{-5} for HIV-gpt viral vector and 5.6×10^{-5} for NL4-3gpt, similar to our results in forward mutation assays. In addition, 6.7×10^{-5} and 2.4×10^{-5} error rates were obtained during the synthesis of the plus strand for both viral vectors, respectively, associated exclusively to the RT.

More recently, a next-generation sequencing approach was conducted to measure the fidelity of transcription and reverse transcription, by using a Single-Molecule Real-Time sequencing (SMRT-Seq) method (**Figure 41**) (Potapov *et al*., 2018). The RNA template was synthesized by the T7 RNAP and reverse transcribed by three different commercial RTs related to those of MLV and AMV. All tested samples showed first-strand error rates ranging from 5.6×10^{-5} to 7.5×10^{-5} (cumulatively made by T7 RNAP and RTs on an RNA template), whereas the second-strand error rates ranged between 5.2×10^{-5} and 8.4×10^{-5} (produced by RTs on DNA template). First strand error rates were similar to ours (**Table 14**), and despite the presence of transcriptional mutations, we demonstrated that the accuracy of RTs belonging to diverse retroviruses such as HIV-1, MLV or AMV was not the same using DNA or RNA templates. These differences in the fidelity of RTs depending on the nature of the template (i.e., RNA or DNA) should be taken into account when comparing error rates obtained in DNA-templated reactions with those obtained in RNA-templated reactions that include transcription errors.

6. CONCLUSIONS

- **Conclusion 1.** WT HIV-2_{ROD} RT showed 1.6-fold lower mutant frequency than HIV-1_{BH10} RT in forward mutation assays based on DNA-dependent DNA synthesis reactions. Meanwhile, K65R in HIV-2_{ROD} RT had almost no effect on the overall fidelity of the enzyme, contrary to the increased intrinsic fidelity observed with this amino acid change in highly divergent HIV-1 RT contexts (group M and O).

- **Conclusion 2.** In HIV-1 RT, drug resistance mutations have been previously associated with an increased fidelity. In contrast, our data revealed that in HIV-2 this relationship cannot be established, particularly for the major mutations found in HIV-2 strains resistant to NRTIs (K65R, Q151M, M184V).

- **Conclusion 3.** Retroviral RTs such as HIV-1_{BH10} (group M subtype B) and HIV-1_{ESP49} (group O) are more faithful when copying RNA than DNA templates. On the other hand, for WT AMV and MLV RTs, and mutant polymerases K65R_{ESP49} and K65R/V75I_{ESP49}, differences between DNA- and RNA-dependent accuracy cannot be determined in a reliable manner due to the sequence heterogeneity of the RNA template.

- **Conclusion 4.** The heterogeneity of RNA templates (generated as a consequence of errors made by T7 RNAP during transcription) limit the correct assessment of fidelity in RNA-dependent DNA synthesis reactions, even after changing the reaction conditions to improve the quality of the RNA template. T7 RNAP synthesizes produced less epimutations in the presence of lower pH and higher magnesium concentrations (i.e., pH 6.75/1.5 mM Mg^{2+} versus pH 7.9/6 mM Mg^{2+}). Despite limitations, HIV-1_{BH10} RT remains as the least accurate RT, and the mutant K65R/V75I_{ESP49} is one of the most faithful RTs.

- **Conclusion 5.** Mutational spectra of retroviral RTs on RNA compared to DNA templates showed differences in hotspot distribution and specificity. Based on the spectra, reverse transcription initiation is revealed as an error-prone process, due to the accumulation of deletion mutations observed in the first two positions following the primer. The hotspot of one-thymidine insertions at position +139 is likely produced by T7 RNAP.

- **Conclusion 6.** Deep sequencing experiments based on the use of Primer IDs allow a broader measurement of RNA-dependent DNA synthesis fidelity of HIV-1 RTs, showing slight differences in fidelity between HIV-1_{BH10} and K65R/V75I_{ESP49} RTs. Although a more detailed analysis of these data is still needed, similar results were obtained with next-generation sequencing experiments in comparison with M13mp2 forward mutation assays. A better assessment of reverse transcription fidelity is expected to be helpful to improve next-generation sequence platforms that require retroviral RTs for RNA sequencing.

- **Conclusión 1.** La RT de tipo salvaje (WT) del VIH-2_{ROD} mostró una frecuencia de mutantes 1,6 veces menor que la del VIH-1_{BH10} en ensayos genéticos basados en reacciones de síntesis de ADN dependientes de ADN. Por otro lado, la RT del VIH-2_{ROD} con el cambio K65R no tuvo prácticamente ningún efecto sobre la fidelidad global de la enzima. Por el contrario, publicaciones previas mostraron una mayor fidelidad intrínseca de K65R, en el contexto de aminoácidos de las RTs de grupos del VIH-1 altamente divergentes (grupos M y O).

- **Conclusión 2.** En la RT del VIH-1, las mutaciones de resistencia a fármacos antirretrovirales se han asociado con una mayor fidelidad (por ejemplo, K65R o M184V). Sin embargo, nuestros datos revelaron que en el VIH-2 no se puede establecer dicha relación, particularmente para las principales mutaciones encontradas en cepas del VIH-2_{ROD} resistentes a análogos de nucleósido (K65R, Q151M, M184V).

- **Conclusión 3.** Las RTs de retrovirus, como las del VIH-1_{BH10} (grupo M subtipo B) y del VIH-1_{ESP49} (grupo O), son más fieles al sintetizar ADN en moldes de ARN que de ADN. Por otro lado, no se pudieron determinar de manera fiable las diferencias de fidelidad de síntesis de ADN en función del molde (ARN o ADN) para las RTs WT del AMV y el MLV, así como para las polimerasas mutantes K65R$_{ESP49}$ y K65R/V75I$_{ESP49}$, debido a la heterogeneidad de la secuencia del molde de ARN, utilizado en las reacciones de retrotranscripción.

- **Conclusión 4.** La heterogeneidad de los moldes de ARN (generada como consecuencia de los errores cometidos por la ARN polimerasa del fago T7 durante la transcripción) limita la determinación correcta de la fidelidad de las RTs en reacciones de síntesis de ADN dependientes de ARN. Esta limitación persiste incluso después de cambiar las condiciones de transcripción para mejorar la calidad del molde de ARN. La ARN polimerasa T7 produce menos errores al sintetizar el ARN en condiciones de menor pH y mayor concentración de magnesio (es decir, pH 6,75/1,5 mM Mg^{2+} frente a pH 7,9/6 mM Mg^{2+}). A pesar de las limitaciones, la RT del VIH-1_{BH10} permanece como la RT menos precisa, y el mutante K65R/V75I$_{ESP49}$ es una de las RTs más fieles.

- **Conclusión 5.** Los espectros de mutación generados por diversas RTs de retrovirus al copiar moldes de ARN, en comparación con aquellos producidos al copiar ADN, mostraron diferencias en la especificidad y distribución de errores. En base a los espectros, el comienzo de la retrotranscripción se revela como una etapa propensa a la introducción de errores, observándose una acumulación de deleciones en las dos primeras posiciones después del iniciador. La acumulación de inserciones de una base de timidina en la posición +139 probablemente sea producida por la ARN polimerasa del fago T7.

- **Conclusión 6.** Los experimentos de secuenciación masiva basados en el uso de iniciadores degenerados permiten una determinación más precisa de la fidelidad de síntesis de ADN dependiente de ARN de las RTs del VIH-1, aunque muestran pequeñas diferencias de fidelidad entre las RTs del VIH-1_{BH10} y del doble mutante K65R/V75I_{ESP49}. Los experimentos basados en tecnologías de secuenciación de nueva generación (NGS) arrojaron resultados similares a los ensayos genéticos basados en el uso del fago M13mp2, si bien es necesario hacer un análisis más detallado de estos datos. Una evaluación más precisa de la fidelidad de la retrotranscripción permitirá mejorar las plataformas de NGS que requieren el uso de RTs de retrovirus para la secuenciación del ARN.

7. REFERENCES

Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. (2010) Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J. Virol*. 84, 9864–78.

Acevedo, A. & Andino, R. (2014) Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc*. 9, 1760−9.

Acevedo, A., Brodsky, L. & Andino, R. (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686−90.

Achuthan, V., Keith, B. J., Connolly, B. A. & DeStefano, J. J. (2014) Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions in vitro. *J. Virol*. 88, 8514–27.

Adamson, C. S. & Freed, E. O. (2007) Human immunodeficiency virus type 1 assembly, release, and maturation. *Adv. Pharmacol.* 55, 347-87.

Álvarez, M., Barrioluengo, V., Afonso-Lehmann, R. N. & Menéndez-Arias, L. (2013) Altered error specificity of RNase H-deficient HIV-1 reverse transcriptases during DNA-dependent DNA synthesis. *Nucleic Acids Res*. 41, 4601–12.

Álvarez, M., Matamoros, T. & Menéndez-Arias, L. (2009) Increased thermostability and fidelity of DNA synthesis of wild-type and mutant HIV-1 group O reverse transcriptases. *J. Mol. Biol.* 392, 872−84.

Álvarez, M., Nevot, M., Mendieta, J., Martínez, M. A. & Menéndez-Arias, L. (2018) Amino acid residues in HIV-2 reverse transcriptase that restricts the development of nucleoside analogue resistance through the excision pathway. *J. Biol. Chem*. 293, 2247−59.

Álvarez, M., Sebastián-Martín, A., García-Marquina, G. & Menéndez-Arias L. (2017) Fidelity of classwide-resistant HIV-2 reverse transcriptase and differential contribution of K65R to the accuracy of HIV-1 and HIV-2 reverse transcriptases. *Sci. Rep*. 7, 44834.

Ambrose, Z. & Aiken, C. (2014) HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* 454-455, 371−9.

Anand, V. S. & Patel, S. S. (2006) Transient state kinetics of transcription elongation by T7 RNA polymerase. *J. Biol. Chem*. 281, 35677–85.

Anderson, J. L. & Hope, T. J. (2004) HIV accessory proteins and surviving the host cell. *Curr. HIV/AIDS Rep.* 1, 47−53.

Andreatta, K., Miller, M. D. & White, K. L. (2013) HIV-2 antiviral potency and selection of drug resistance mutations by the integrase strand transfer inhibitor elvitegravir and NRTIs emtricitabine and tenofovir in vitro. *J. Acquir. Immune Defic. Syndr.* 62, 367–74.

Arezi, B. & Hogrefe, H. (2009) Novel mutations in Moloney murine leukemia virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* 37, 473–81.

Arion, D., Kaushik, N., McCormick, S., Borkow, G. & Parniak, M. A. (1998) Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry* 37, 15908−17.

Back, N. K. T., Nijhuis, M., Keulen, W., Boucher, C. A., Oude Essink, B. O., van Kuilenburg, A. B., van Gennip, A. H. & Berkhout, B. (1996) Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme. *EMBO J.* 15, 4040–49.

Bakhanashvili, M. & Hizi, A. (1992a) Fidelity of the RNA-dependent DNA synthesis exhibited by the reverse transcriptases of human immunodeficiency virus types 1 and 2 and of murine leukemia virus: mispair extension frequencies. *Biochemistry* 31, 9393–8.

Bakhanashvili, M. & Hizi, A. (1992b) Fidelity of the reverse transcriptase of human immunodeficiency virus type 2. *FEBS Lett.* 306, 151–6.

Bakhanashvili, M. & Hizi, A. (1993) The fidelity of the reverse transcriptases of human immunodeficiency viruses and murine leukemia virus, exhibited by the mispair extension frequencies, is sequence dependent and enzyme related. *FEBS Lett*. 319, 201–5.

Baranauskas, A., Paliksa, S., Alzbutas, G., Vaitkevicius, M., Lubiene, J., Letukiene, V., Burinskas, S., Sasnauskas, G. & Skirgaila, R. (2012) Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng. Des. Sel*. 25, 657–68.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868-71.

Barrioluengo, V. (2013) Diseño, purificación y caracterización de variantes de la retrotranscriptasa del virus de la inmunodeficiencia humana de interés biotecnológico (tesis doctoral). Universidad Autónoma de Madrid, Madrid (España). Available at the repository: https://repositorio.uam.es/handle/10486/660191.

Barrioluengo, V., Álvarez, M., Barbieri, D. & Menéndez-Arias, L. (2011) Thermostable HIV-1 group O reverse transcriptase variants with the same fidelity as murine leukaemia virus reverse transcriptase. *Biochem. J.* 436, 599–607.

Barrioluengo, V., Wang, Y., Le Grice, S. F. & Menéndez-Arias L. (2012) Intrinsic DNA synthesis fidelity of xenotropic murine leukemia virus-related virus reverse transcriptase. *FEBS J.* 279, 433–44.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantittes of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. 72, 248–54.

Bebenek, K., Roberts, J. D. & Kunkel, T. A. (1992) The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *J. Biol. Chem.* 267, 3589–96.

Bebenek, K. & Kunkel, T. A. (1995) Analyzing fidelity of DNA polymerases. *Methods Enzymol*. 262, 217–32.

Beckman, R. A., Mildvan, A. S. & Loeb, L. A. (1985) On the fidelity of DNA replication: manganese mutagenesis in vitro. *Biochemistry* 24, 5810–7.

Bird, L. E., Chamberlain, P. P., Stewart-Jones, G. B., Ren, J., Stuart, D. I. & Stammers, D. K. (2003). Cloning, expression, purification, and crystallisation of HIV-2 reverse transcriptase. *Protein Expr. Purif.* 27, 12–18.

Bohlayer, W. P. & DeStefano, J. J. (2006) Tighter binding of HIV reverse transcriptase to RNA-DNA versus DNA-DNA results mostly from interactions in the polymerase domain and requires just a small stretch of RNA-DNA. *Biochemistry* 45, 7628–38.

Boretto, J., Longhi, S., Navarro, J. M., Selmi, B., Sire, J. & Canard, B. (2001) An integrated system to study multiply substituted human immunodeficiency virus type 1 reverse transcriptase. *Anal. Biochem.* 292, 139–47.

Boyer, J. C., Bebenek, K. & Kunkel, T. A. (1992) Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. U.S.A.* 89, 6919–23.

Boyer, P. L., Clark, P. K. & Hughes, S. H. (2012) HIV-1 and HIV-2 reverse transcriptases: different mechanisms of resistance to nucleoside reverse transcriptase inhibitors. *J. Virol.* 86, 5885–94.

Boyer, P.L. & Hughes, S. H. (2000) Effects of amino acid substitutions at position 115 on the fidelity of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* 74, 6494-500.

Boyer, P. L., Sarafianos, S. G., Clark, P. K., Arnold, E. & Hughes, S. H. (2006) Why do HIV-1 and HIV-2 use different pathways to develop AZT resistance? *PLoS Pathog*. 2, e10.

Boyer, P. L., Stenbak, C. R., Hoberman, D., Linial, M. L. & Hughes, S. H. (2007) In vitro fidelity of the prototype primate foamy virus (PFV) RT compared to HIV-1 RT. *Virology* 367, 253–64.

Brancaccio, R. N., Robitaille, A., Dutta, S., Cuenin, C., Santare, D., Skenders, G., Leja, M., Fischer, N., Giuliano, A. R., Rollison, D. E., Grundhoff, A., Tommasino, M. & Gheit, T. (AÑO) Generation of a novel next-generation sequencing-based method for the isolation of new human papillomavirus types. *Virology* 520, 1–10.

Brandin, E., Lindborg, L., Gyllensten, K., Broström, C., Hagberg, L., Gisslen, M., Tuvesson, B., Blaxhult, A. & Albert, J. (2003) Pol gene sequence variation in Swedish HIV-2 patients failing antiretroviral therapy. *AIDS Res. Hum. Retroviruses* 19, 543–50.

Briggs, J. A. & Kräusslich, H. G. (2011) The molecular architecture of HIV. *J. Mol. Biol*. 410, 491–500.

Brosius, S., Grosse, F. & Krauss, G. (1983) Subspecies of DNA polymerase alpha from calf thymus with different fidelity in copying synthetic template-primers. *Nucleic Acids Res.* 11, 193–202.

Bubunenko, M. G., Court, C. B., Rattray, A. J., Gotte, D. R., Kireeva, M. L., Irizarry-Caro, J. A., Li, X., Jin, D. J., Court, D. L., Strathern, J. N. & Kashlev, M. (2017) A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in Escherichia coli. *Genetics* 206, 179−87.

Canard, B., Chowdhury, K., Sarfati, R., Doublié, S. & Richardson, C. C. (1999) The motif D loop of human immunodeficiency virus type 1 reverse transcriptase is critical for nucleoside 5'-triphosphate selectivity. *J. Biol. Chem.* 274, 35768–76.

Capobianchi, M. R., Giombini, E. & Rozera, G. (2012) Next-generation sequencing technology in clinical virology. *Clin. Microbiol. Infect.* 19, 15–22.

Cary, D. C., Fujinaga, K. & Peterlin, B. M. (2016) Molecular mechanisms of HIV latency. *J. Clin. Invest.* 126, 448–54.

Cases-González, C. E., Gutiérrez-Rivas, M. & Menéndez-Arias, L. (2000) Coupling ribose selection to fidelity of DNA synthesis: the role of Tyr-115 of human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem*. 275, 19759–67.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. & Debyser, Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278, 372–81.

Chung, S., Miller, J. T., Lapkouski, M., Tian, L., Yang, W. & Le Grice, S. F. (2013) Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. *J. Biol. Chem.* 288, 16177–84.

Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M. A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L. & Montagnier, L. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science* 233, 343–6.

Clavel, F., Guyader, M., Guétard, D., Sallé, M., Montagnier, L. & Alizon, M. (1986) Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* 324, 691–5.

Cohen, E. A., Dehni, G., Sodroski, J. G. & Haseltine, W. A. (1990) Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J. Virol.* 64, 3097–9.

Cong, M., Heneine, W. & García-Lerma, J. G. (2007) The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. *J. Virol*. 81, 3037–41.

D'arc, M., Ayouba, A., Esteban, A., Learn, G. H., Boué, V., Liegeois, F., Etienne, L., Tagg, N., Leendertz, F. H., Boesch, C., Madinda, N. F., Robbins, M. M., Gray, M., Cournil, A., Ooms, M., Letko, M., Simon, V. A., Sharp, P. M., Hahn, B. H., Delaporte, E., Mpoudi Ngole, E. & Peeters, M. (2015) Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1343–52.

Damond, F., Collin, G., Matheron, S., Peytavin, G., Campa, P., Delarue, S., Taieb, A., Bénard, A., Chêne, G., Brun-Vézinet, F., Descamps, D. & French ANRS HIV-2 Cohort (ANRS CO 5 VIH-2). (2005) Letter. In vitro phenotypic susceptibility to nucleoside reverse transcriptase inhibitors of HIV-2 isolates with the Q151M mutation in the reverse transcriptase gene. *Antivir. Ther.* 10, 861–5.

Das, K., Bandwar, R. P., White, K. L., Feng, J. Y., Sarafianos, S. G., Tuske, S., Tu, X., Clark, A. D. Jr., Boyer, P. L., Hou, X., Gaffney, B. L., Jones, R. A., Miller, M. D., Hughes, S. H. & Arnold, E. (2009) Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *J. Biol. Chem*. 284, 35092−100.

Das, K., Martinez, S. E., Bandwar, R. P. & Arnold, E. (2014) Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. *Nucleic Acids Res*. 42, 8125–37.

Das, K., Martinez, S. E. & Arnold, E. (2017) Structural insights into HIV reverse transcriptase mutations Q151M and Q151M complex that confer multinucleoside drug resistance. *Antimicrob. Agents Chemother*. 61, e00224−17.

Darlix, J. L., Godet, J., Ivanyi-Nagy, R., Fossé, P., Mauffret, O. & Mély, Y. (2011) Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol*. 410, 565–81.

Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 81, 2035–9.

de Paz, A. M., Cybulski, T. R., Marblestone, A. H., Zamft, B. M., Church, G. M., Boyden, E. S., Kording, K. P. & Tyo, K. E. J. (2018) High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Res*. 46, e78.

Delviks-Frankenberry, K. A., Nikolenko, G. N. & Pathak, V. K. (2010) The "connection" between HIV drug resistance and RNase H. *Viruses* 2, 1476−503.

Descamps, D., Damond, F., Matheron, S., Collin, G., Campa, P., Delarue, S., Pueyo, S., Chêne, G., Brun-Vézinet, F. & French ANRS HIV-2 Cohort Study Group. (2004) High frequency of selection of K65R and Q151M mutations in HIV-2 infected patients receiving nucleoside reverse transcriptase inhibitors containing regimen. *J. Med. Virol*. 74, 197−201.

Deuzing, I. P., Charpentier, C., Wright, D. W., Matheron, S., Paton, J., Frentz, D., van de Vijver, D. A., Coveney, P. V., Descamps, D., ANRS CO5 HIV-2 Cohort, Boucher, C. A. & Beerens, N. (2015) Mutation V111I in HIV-2 reverse transcriptase increases the fitness of the nucleoside analogue-resistant K65R and Q151M viruses. *J. Virol.* 89, 833–43.

Deval, J., White, K. L., Miller, M. D., Parkin, N. T., Courcambeck, J., Halfon, P., Selmi, B., Boretto, J. & Canard, B. (2004) Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. *J. Biol. Chem*. 279, 509–16.

Deval, J., Alvarez, K., Selmi, B., Bermond, M., Boretto, J., Guerreiro, C., Mulard, L. & Canard, B. (2005) Mechanistic insights into the suppression of drug resistance by human immunodeficiency virus type 1 reverse transcriptase using α-boranophosphate nucleoside analogs. *J. Biol. Chem*. 280, 3838–46.

di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1986) Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science* 231, 1289–91.

Diamond, T. L., Souroullas, G., Weiss, K. K., Lee, K. Y., Bambara, R. A., Dewhurst, S. & Kim, B. (2003) Mechanistic understanding of an altered fidelity simian immunodeficiency virus reverse transcriptase mutation, V148I, identified in a pig-tailed macaque. *J. Biol. Chem*. 278, 29913–24.

Divita, G., Rittinger, K., Restle, T., Immendorfer, U. & Goody, R. S. (1995). Conformational stability of dimeric HIV-1 and HIV-2 reverse transcriptases. *Biochemistry* 34, 16337–46.

Dubois, N., Marquet, R., Paillart, J. C. & Bernacchi, S. (2018) Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.* 9, 527.

Dufour, E., El Dirani-Diab, R., Boulmé, F., Fournier, M., Nevinsky, G., Tarrago-Litvak, L., Litvak, S. & Andreola, M. L. (1998) p66/p51 and p51/p51 recombinant forms of reverse transcriptase from human immunodeficiency virus type 1. Interactions with primer $tRNA^{Lys3}$, initiation of cDNA synthesis, and effect of inhibitors. *Eur. J. Biochem*. 251, 487–95.

Echols, H. & Goodman, M. F. (1991) Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem*. 60, 477–511.

Eckert, K. A. & Kunkel, T. A. (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res*. 18, 3739–44.

Eckert, K. A. & Kunkel, T. A. (1993a) Effect of reaction pH on the fidelity and processivity of exonuclease-deficient Klenow polymerase. *J. Biol. Chem*. 268, 13462–71.

Eckert, K. A. & Kunkel, T. A. (1993b) Fidelity of DNA synthesis catalyzed by human DNA polymerase α and HIV-1 reverse transcriptase: effect of reaction pH. *Nucleic Acids Res*. 21, 5212–20.

Ellinger, T. & Ehricht, R. (1998) Single-step purification of T7 RNA polymerase with a 6-histidine tag. *Biotechniques*. 24, 718–20. Erratum in: 25, 640 (1998).

Esposito, F., Corona, A. & Tramontano, E. (2012) HIV-1 reverse transcriptase still remains a new drug target: structure, function, classical inhibitors, and new inhibitors with innovative mechanisms of actions. *Mol. Biol. Int*. 2012, 586401.

Fan, N., Rank, K. B., Leone, J. W., Heinrikson, R. L., Bannow, C. A., Smith, C. W., Evans, D. B., Poppe, S. M., Tarpley, W. G., Rothrock, D. J., Tomasselli, A. G. & Sharma, S. K. (1995). The differential processing of homodimers of reverse transcriptases from human immunodeficiency viruses type 1 and 2 is a consequence of the distinct specificities of the viral proteases. *J. Biol. Chem.* 270, 13573–9.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G. & Lemey, P. (2014) HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346, 56–61.

Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1infection. *Virus Res*. 170, 15–24.

Faust, T. B., Binning, J. M., Gross, J. D. & Frankel, A. D. (2017) Making sense of multifunctional proteins: human immunodeficiency virus type 1 accessory and regulatory proteins and connections to transcription. *Annu. Rev. Virol*. 4, 241–26.

Fisher, T. S., Darden, T. & Prasad, V. R. (2003) Substitutions at Phe61 in the β3-β4 hairpin of HIV-1 reverse transcriptase reveal a role for the fingers subdomain in strand displacement DNA synthesis. *J. Mol. Biol*. 325, 443–59.

Foley, B., Leitner, T., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., Wolinsky, S. & Korber, B. (Eds.) (2018) HIV sequence compendium 2018. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 18-25673.

Folks, T. M., Powell, D., Lightfoote, M., Koenig, S., Fauci, A. S., Benn, S., Rabson, A., Daugherty, D., Gendelman, H. E., Hoggan, M. D., Venkatesan, S. & Martin, M. A. (1986). Biological and biochemical characterization of a cloned Leu-3^{-} cell surviving infection with the acquired immune deficiency syndrome retrovirus. *J. Exp. Med.* 164, 280–90.

Francis, A. C. & Melikyan, G. B. (2018a) Single HIV-1 imaging reveals progression of infection through CA-dependent steps of docking at the nuclear pore, uncoating, and nuclear transport. *Cell Host Microbe* 23, 536–48.

Francis, A. C. & Melikyan, G. B. (2018b) Live-cell imaging of early steps of single HIV-1 infection. *Viruses* 10, E275.

Freed, E. O. (2015) HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* 13, 484–96.

Gallego, I., Gregori, J., Soria, M. E., García-Crespo, C., García-Álvarez, M., Gómez-González, A., Valiergue, R., Gómez, J., Esteban, J. I., Quer, J., Domingo, E. & Perales, C. (2018) Resistance of high fitness hepatitis C virus to lethal mutagenesis. *Virology* 523, 100–9.

Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P. & Markham, P. D. (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 224, 500–3.

Gamba, P. & Zenkin, N. (2018) Transcription fidelity and its roles in the cell. *Curr. Opin. Microbiol*. 42, 13−8.

Ganser-Pornillos, B. K., Yeager, M. & Pornillos, O. (2012) Assembly and architecture of HIV. *Adv. Exp. Med. Biol*. 726, 441–65.

Golinelli, M. P. & Hughes, S. H. (2002). Nontemplated nucleotide addition by HIV-1 reverse transcriptase. *Biochemistry* 41, 5894–906.

González, M. E. (2017) The HIV-1 Vpr protein: a multifaceted target for therapeutic intervention. *Int. J. Mol. Sci.* 18, E126.

Gordon, A. J., Satory, D., Halliday, J. A. & Herman, C. (2015) Lost in transcription: transient errors in information transfer. *Curr. Opin. Microbiol*. 24, 80–7.

Gottlieb, G. S., Badiane, N. M., Hawes, S. E., Fortes, L., Toure, M., Ndour, C. T., Starling, A. K., Traore, F., Sall, F., Wong, K. G., Cherne, S. L., Anderson, D. J., Dye, S. A., Smith, R. A., Mullins, J. I., Kiviat, N. B., Sow, P. S. & University of Washington-Dakar HIV-2 Study Group. (2009) Emergence of multiclass drug-resistance in HIV-2 in antiretroviral-treated individuals in Senegal: implications for HIV-2 treatment in resource-limited West Africa. *Clin. Infect. Dis*. 48, 476–83. Erratum in: 48, 848 (2009).

Gout, J. F., Thomas, W. K., Smith, Z., Okamoto, K. & Lynch, M. (2013) Large-scale detection of in vivo transcription errors. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18584–9.

Gout, J. F., Li, W., Fritsch, C., Li, A., Haroon, S., Singh, L., Hua, D., Fazelinia, H., Smith, Z., Seeholzer, S., Thomas, K., Lynch, M. & Vermulst, M. (2017) The landscape of transcription errors in eukaryotic cells. *Sci. Adv.* 3, e1701484.

Günthard, H. F., Saag, M. S., Benson, C. A., del Rio, C., Eron, J. J., Gallant, J. E., Hoy, J. F., Mugavero, M. J., Sax, P. E., Thompson, M. A., Gandhi, R. T., Landovitz, R. J., Smith, D. M.,

Jacobsen, D. M. & Volberding, P. A. (2016) Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2016 recommendations of the International Antiviral Society-USA Panel. *JAMA* 316, 191–210.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* 326, 662–9.

Hachiya, A., Kodama, E. N., Schuckmann, M. M., Kirby, K. A., Michailidis, E., Sakagami, Y., Oka, S., Singh, K. & Sarafianos, S. G. (2011) K70Q adds high-level tenofovir resistance to "Q151M complex" HIV reverse transcriptase through the enhanced discrimination mechanism. *PLoS ONE* 6, e16242.

Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. (2000) Role of murine leukemia virus reverse transcriptase deoxyribonucleoside triphosphate-binding site in retroviral replication and in vivo fidelity. *J. Virol*. 74, 10349–58.

Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Janzen, D. H., Hallwachs, W., Naik, S., Sones, J. E. & Zakharov, E. V. (2018) A Sequel to Sanger: amplicon sequencing that scales. *BMC Genomics* 19, 219.

Herschhorn, A. & Hizi, A. (2010) Retroviral reverse transcriptases. *Cell Mol. Life Sci*. 67, 2717–47.

Herzig, E. & Hizi, A. (2015) The importance of glutamine 294 that affects the ribonuclease H activity of the reverse transcriptase of HIV-2 to viral replication. *Virology* 483, 13–20.

Hidalgo, L. & Swanson, C. M. (2017) Regulation of human immunodeficiency virus type 1 (HIV-1) mRNA translation. *Biochem. Soc. Trans*. 45, 353–64.

Hostomsky, Z., Hostomska, Z., Fu, T. B. & Taylor, J. (1992) Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. *J. Virol.* 66, 3179–82.

Houbiers, M. C. & Hemminga, M. A. (2004) Protein-lipid interactions of bacteriophage M13 gene 9 minor coat protein. *Mol. Membr. Biol.* 21, 351–9.

Hu, W. S. & Hughes, S. H. (2012) HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med*. 2, a006882.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282, 1669–75.

Huang, J., Brieba, L. G. & Sousa, R. (2000) Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* 39, 11571–80.

Hübner, A., Kruhoffer, M., Grosse, F. & Krauss, G. (1992) Fidelity of human immunodeficiency virus type I reverse transcriptase in copying natural RNA. *J. Mol. Biol.* 223, 595–600.

Hulme, A. E., Perez, O. & Hope, T. J. (2011) Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9975–80.

Ikeda, R. A. & Richardson, C. C. (1987) Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. *J. Biol. Chem.* 262, 3790−9.

Imashimizu, M., Oshima, T., Lubkowska, L. & Kashlev, M. (2013) Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res*. 41, 9090–104.

Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. & Swanstrom, R. (2011) Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20166–71.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6320–4.

Jacques, D. A., McEwan, W. A., Hilditch, L., Price, A. J., Towers, G. J. & James, L. C. (2016) HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* 536, 349–53.

Ji, J. & Loeb, L. A. (1992) Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* 31, 954–8.

Ji, J. & Loeb, L. A. (1994) Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 env gene. *Virology* 199, 323–30.

Johnson, K. A. (1993) Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* 62, 685–713.

Kao, S., Akari, H., Khan, M. A., Dettenhofer, M., Yu, X. F. & Strebel, K. (2003) Human immunodeficiency virus type 1 Vif is efficiently packaged into virions during productive but not chronic infection. *J. Virol.* 77, 1131–40.

Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. (1992) Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* 267, 25988–97.

Kehoe, J. W. & Kay, B. K. (2005) Filamentous phage display in the new millennium. *Chem. Rev.* 105, 4056–72.

Kennedy, S. R., Schmitt, M. W., Fox, E. J., Kohrn, B. F., Salk, J. J., Ahn, E. H., Prindle, M. J., Kuong, K. J., Shen, J. C., Risques, R. A. & Loeb, L. A. (2014) Detecting ultralow-frequency mutations by duplex sequencing. *Nat. Protoc.* 9, 2586–606.

Kerr, S. G. & Anderson, K. S. (1997) RNA dependent DNA replication fidelity of HIV-1 reverse transcriptase: evidence of discrimination between DNA and RNA substrates. *Biochemistry* 36, 14056–63.

Keulen, W., Back, N. K., van Wijk, A., Boucher, C. A. & Berkhout, B. (1997) Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol*. 71, 3346–50.

Kielpinski, L. J., Hagedorn, P. H., Lindow, M., Vinther, J. (2017) RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic Acids Res.* 45, 12932–44.

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9530–5.

King, G. C., Martin, C. T., Pham, T. T. & Coleman, J. E. (1986) Transcription by T7 RNA polymerase is not zinc-dependent and is abolished on amidomethylation of cysteine-347. *Biochemistry* 25, 36–40.

Knippa, K. & Peterson, D. O. (2013) Fidelity of RNA polymerase II transcription: Role of Rpb9 in error detection and proofreading. *Biochemistry* 52, 7807−17. Erratum in: 54, 6294 (2015).

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256, 1783–90.

Krupovic, M., Blomberg, J., Coffin, J. M., Dasgupta, I., Fan, H., Geering, A. D., Gifford, R., Harrach, B., Hull, R., Johnson, W., Kreuze, J. F., Lindemann, D., Llorens, C., Lockhart, B., Mayer, J., Muller, E., Olszewski, N. E., Pappu, H. R., Pooggin, M. M., Richert-Pöggeler, K. R., Sabanadzovic, S., Sanfaçon, H., Schoelz, J. E., Seal, S., Stavolone, L., Stoye, J. P., Teycheney, P. Y., Tristem, M., Koonin, E. V. & Kuhn, J. H. (2018) Ortervirales: New virus order unifying five families of reverse-transcribing viruses. *J. Virol.* 92, e00515-18.

Kunkel, T. A. & Soni, A. (1988) Mutagenesis by transient misalignment. *J. Biol. Chem*. 263, 14784–9.

Laguette, N., Rahm, N., Sobhian, B., Chable-Bessia, C., Münch, J., Snoeck, J., Sauter, D., Switzer, W. M., Heneine, W., Kirchhoff, F., Delsuc, F., Telenti, A. & Benkirane, M. (2012) Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe* 11, 205–17.

Lapkouski, M., Tian, L., Miller, J. T., Le Grice, S. F. J. & Yang, W. (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol*. 20, 230–6.

Larder, B. A. (1994) Interactions between drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase. *J. Gen. Virol*. 75, 951–57.

Larder, B. A., Purifoy, D. J., Powell, K. L. & Darby, G. (1987) Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature* 327, 716–7.

Le Grice, S. F., Naas, T., Wohlgensinger, B., Schatz, O. (1991) Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J.* 10, 3905–11.

Lemey, P., Kosakovsky Pond, S. L., Drummond, A. J., Pybus, O. G., Shapiro, B., Barroso, H., Taveira, N. & Rambaut, A. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput. Biol*. 3, e29.

Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M. & Oshiro, L. S. (1984) Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 225, 840–2.

Lim, E. S., Fregoso, O., McCoy, C. O., Matsen, F. A., Malik, H. S. & Emerman, M. (2012) The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* 11, 194–204.

Lis, J. T. (1980) Fractionation of DNA fragments by polyethylene glycol induced precipitation. *Methods Enzymol.* 65, 347–353.

Lloyd, S. B., Lloyd, S. B., Lichtfuss, M., Amarasena, T. H., Alcantara, S., De Rose, R., Tachedjian, G., Alinejad-Rokny, H., Venturi, V., Davenport, M. P., Winnall, W. R. & Kent, S. J. (2016) High fidelity simian immunodeficiency virus reverse transcriptase mutants have impaired replication in vitro and in vivo. *Virology* 492, 1–10.

Lowe, D. M., Parmar, V., Kemp, S. D. & Larder, B. A. (1991) Mutational analysis of two conserved sequence motifs in HIV-1 reverse transcriptase. *FEBS Lett*. 282, 231–4.

Luczkowiak, J., Matamoros, T. & Menéndez-Arias, L. (2018) Template-primer binding affinity and RNase H cleavage specificity contribute to the strand transfer efficiency of HIV-1 reverse transcriptase. *J. Biol. Chem.* 293, 13351–63.

MacNeil, A., Sankale, J. L., Meloni, S. T., Sarr, A. D., Mboup, S. & Kanki, P. (2007) Long-term intrapatient viral evolution during HIV-2 infection. *J. Infect. Dis*. 195, 726–33.

Magnuson, B., Bedi, K. & Ljungman, M. (2016) Genome stability versus transcript diversity. *DNA Repair (Amst.)* 44, 81–6.

Majumdar, C., Abbotts, J., Broder, S. & Wilson, S. H. (1998) Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition. *J. Biol. Chem.* 263, 15657–65.

Mamede, J. I., Cianci, G. C., Anderson, M. R. & Hope, T. J. (2017) Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7169-78.

Mansky, L. M., Le Rouzic, E., Benichou, S. & Gajary, L. C. (2003) Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. *J. Virol*. 77, 2071–80.

Matamoros, T., Barrioluengo, V., Abia, D. & Menéndez-Arias, L. (2013) Major groove binding track residues of the connection subdomain of human immunodeficiency virus type 1 reverse transcriptase enhance cDNA synthesis at high temperatures. *Biochemistry* 52, 9318–28.

Matamoros, T., Deval, J., Guerreiro, C., Mulard, L., Canard, B. & Menéndez-Arias, L. (2005) Suppression of multidrug-resistant HIV-1 reverse transcriptase primer unblocking activity by α-phosphate-modified thymidine analogues. *J. Mol. Biol*. 349, 451–63.

Megens, S. & Laethem, K. V. (2013) HIV-1 genetic variation and drug resistance development. *Expert Rev. Anti Infect. Ther*. 11, 1159–78.

Mellenius, H. & Ehrenberg, M. (2017) Transcriptional accuracy modeling suggests two-step proofreading by RNA polymerase. *Nucleic Acids Res*. 45, 11582−93.

Mendieta, J., Cases-González, C. E., Matamoros, T., Ramírez, G. & Menéndez-Arias, L. (2008) A Mg^{2+}-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins* 71, 565–74.

Menéndez-Arias, L. (2002) Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. *Prog. Nucleic Acid Res. Mol. Biol*. 71, 91–147.

Menéndez-Arias, L. (2008) Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. *Virus Res*. 134, 124–46.

Menéndez-Arias, L. (2009) Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. *Viruses* 1, 1137–65.

Menéndez-Arias, L. (2013a) HIV reverse transcriptase fidelity, clade diversity, and acquisition of drug resistance. In: Human immunodeficiency virus reverse transcriptase: a bench-to-bedside success (ed. Le Grice, S. F. J. & Götte, M.) 225–52 (Springer Science+ Business Media, 2013).

Menéndez-Arias, L. (2013b) Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Res.* 98, 93–120.

Menéndez-Arias, L. & Álvarez, M. (2014) Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antiviral Res*. 102, 70–86.

Menéndez-Arias, L., Sebastián-Martín, A. & Álvarez M. (2017) Viral reverse transcriptases. *Virus Res*. 234, 153–76.

Menéndez-Arias, L. & Tözsér, J. (2008) HIV-1 protease inhibitors: effects on HIV-2 replication and resistance. *Trends Pharmacol. Sci*. 29, 42–9.

Meyer, P. R., Matsuura, S. E., Mian, A. M., So, A. G. & Scott, W. A. (1999) A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. *Mol. Cell* 4, 35–43.

Mishanina, T. V., Palo, M. Z., Nayak, D., Mooney, R. A. & Landick, R. (2017) Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci. U. S. A.* 114, E5103−12.

Mizrahi, V., Usdin, M. T., Harington, A. & Dudding, L. R. (1990) Site-directed mutagenesis of the conserved Asp-443 and Asp-498 carboxy-terminal residues of HIV-1 reverse transcriptase. *Nucleic Acids Res*. 18, 5359– 63.

Mizrahi, V., Brooksbank, R. L. & Nkabinde, N. C. (1994) Mutagenesis of the conserved aspartic acid 443, glutamic acid 478, asparagine 494, and aspartic acid 498 residues in the ribonuclease H domain of p66/p51 human immunodeficiency virus type I reverse transcriptase. Expression and biochemical analysis. *J. Biol. Chem.* 269, 19245–9.

Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V. R., Hunicke-Smith, S., Swamy, S., Kuersten, S. & Lambowitz, A. M. (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* 19, 958–70.

Müller, B., Tessmer, U., Schubert, U. & Kräusslich, H. G. (2000) Human immunodeficiency virus type 1 Vpr protein is incorporated into the virion in significantly smaller amounts than gag and is phosphorylated in infected cells. *J. Virol.* 74, 9727–31.

Ndung'u, T. & Weiss, R. A. (2012) On HIV diversity. *AIDS* 26, 1255–60.

Ne, E., Palstra, R. J. & Mahmoudi, T. (2018) Transcription: insights from the HIV-1 promoter. *Int. Rev. Cell Mol. Biol*. 335, 191–243.

Ninio, J. (1991) Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics* 129, 957–62.

Ntemgwa, M. L., d'Aquin Toni, T., Brenner, B. G., Camacho, R. J. & Wainberg, M. A. (2009) Antiretroviral drug resistance in human immunodeficiency virus type 2. *Antimicrob. Agents Chemother.* 53, 3611–9.

O'Neil, P. K., Sun, G., Yu, H., Ron, Y., Dougherty, J. P. & Preston, B. D. (2002) Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J. Biol. Chem.* 277, 38053−61.

Ohta, T., Tokishita, S., Tsunoi, R., Ohmae, S. & Yamagata, H. (2002) Characterization of Trp(+) reversions in Escherichia coli strain WP2uvrA. *Mutagenesis* 17, 313–6.

Okada, A. & Iwatani, Y. (2016) APOBEC3G-mediated G-to-A hypermutation of the HIV-1 genome: the missing link in antiviral molecular mechanisms. *Front. Microbiol.*7, 2027.

Okano, H., Baba, M., Hidese, R., Iida, K., Li, T., Kojima, K., Takita, T., Yanagihara, I., Fujiwara, S. & Yasukawa, K. (2018) Accurate fidelity analysis of the reverse transcriptase by a modified next-generation sequencing. *Enzyme Microb. Technol.* 115, 81−5.

Operario, D. J., Reynolds, H. M. & Kim, B. (2005) Comparison of DNA polymerase activities between recombinant feline immunodeficiency and leukemia virus reverse transcriptases. *Virology* 335, 106–21.

Orias, E. & Gartner, T. K. (1966) Suppression of amber and ochre rII mutants of bacteriophage T4 by streptomycin. *J. Bacteriol*. 91, 2210–5.

Ozsolak, F. & Milos, P. M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98.

Patel, P. H. & Preston, B. D. (1994) Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. *Proc. Natl. Acad. Sci. U. S. A.* 91, 549–53.

Pennington, M. R., Grenier, J. K. & Van de Walle, G. R. Transcriptome profiling of alphaherpesvirus-infected cells treated with the HIV-integrase inhibitor raltegravir reveals profound and specific alterations in host transcription. *J. Gen. Virol*. 99, 1115–28.

Perrino, F. W., Preston, B. D., Sandell, L. L. & Loeb, L. A. (1989) Extension of mismatched 3' termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* 86, 8343–7.

Peterson, K., Jallow, S., Rowland-Jones, S. L. & de Silva, T. I. (2011) Antiretroviral therapy for HIV-2 infection: recommendations for management in low-resource settings. *AIDS Res. Treat*. 2011, 463704.

Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 224, 497–500.

Potapov, V., Fu, X., Dai, N., Corrêa, I. R. Jr., Tanner, N. A. & Ong, J. L. (2018) Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res.* 46, 5753−63.

Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) Fidelity of HIV-1 reverse transcriptase. *Science* 242, 1168–71.

Quiñones-Mateu, M. E., Albright, J. L., Mas, A., Soriano, V. & Arts, E. J. (1998) Analysis of pol gene heterogeneity, viral quasispecies, and drug resistance in individuals infected with group O strains of human immunodeficiency virus type 1. *J. Virol.* 72, 9002–15.

Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C., Rousso, I. (2017) Reverse transcription mechanically initiates HIV-1 capsid disassembly. *J. Virol*. 91, e00289–17.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R. Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313, 277–84.

Rawson, J. M., Landman, S. R., Reilly, C. S. & Mansky, L. M. (2015) HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology* 12, 60.

Ray, A. S., Fordyce, M. W. & Hitchcock, M. J. (2016). Tenofovir alafenamide: a novel prodrug of tenofovir for the treatment of human immunodeficiency virus. *Antiviral Res*. 125, 63–70.

Reid, P., MacInnes, H., Cong, M. E., Heneine, W. & García-Lerma, J. G. (2005) Natural resistance of human immunodeficiency virus type 2 to zidovudine. *Virology* 336, 251–64.

Reid-Bayliss, K. S. & Loeb, L. A. (2017) Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proc. Natl. Acad. Sci. U. S. A.* 114, 9415–20.

Remington, K. M., Bennett, S. E., Harris, C. M., Harris, T. M. & Bebenek, K. (1998) Highly mutagenic bypass synthesis by T7 RNA polymerase of site-specific benzo[a]pyrene diol epoxide-adducted template DNA. *J. Biol. Chem.* 273, 13170–6.

Ren, J., Bird, L. E., Chamberlain, P. P., Stewart-Jones, G. B., Stuart, D. I. & Stammers, D. K. (2002). Structure of HIV-2 reverse transcriptase at 2.35-Å resolution and the mechanism of resistance to non-nucleoside inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14410–5.

Ribeiro, A. C., Maia e Silva, A., Santa-Marta, M., Pombo, A., Moniz-Pereira, J., Goncalves, J. & Barahona, I. (2005) Functional analysis of Vif protein shows less restriction of human immunodeficiency virus type 2 by APOBEC3G. *J. Virol.* 79, 823–33.

Richter, S. N., Frasson, I. & Palù, G. (2009) Strategies for inhibiting function of HIV-1 accessory proteins: a necessary route to AIDS therapy? *Curr. Med. Chem*. 16, 267–86.

Roberts, J. D., Bebenek, K. & Kunkel, T. A. (1988) The accuracy of reverse transcriptase from HIV-1. *Science* 242, 1171–3.

Roberts, J. D., Preston, B. D., Johnston, L. A., Soni, A., Loeb, L. A. & Kunkel, T. A. (1989) Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Mol. Cell Biol*. 9, 469–76.

Romani, B. & Engelbrecht, S. (2009) Human immunodeficiency virus type 1 Vpr: functions and molecular interactions. *J. Gen. Virol*. 90, 1795–805.

Saleh, S., Vranckx, L., Gijsbers, R., Christ, F. & Debyser, Z. (2017) Insight into HIV-2 latency may disclose strategies for a cure for HIV-1 infection. *J. Virus Erad*. 3, 7–14.

Sanger, F. & Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–8.

Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D. Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H. & Arnold, E. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* 20, 1449−61.

Sarafianos, S. G., Marchand, B., Das, K., Himmel, D., Parniak, M. A., Hughes, S. H. & Arnold, E. (2009) Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol*. 385, 693–713.

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B. & Loeb, L. A. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14508–13.

Schultz, S. J. & Champoux, J. J. (2008) RNase H activity: structure, specificity, and function in reverse transcription. *Virus Res*. 134, 86–103.

Schultz, S. J., Zhang, M. & Champoux, J. J. (2009) Preferred sequences within a defined cleavage window specify DNA 3' end-directed cleavages by retroviral RNases H. *J. Biol. Chem*. 284, 32225–38.

Sevilya, Z., Loya, S., Hughes, S. H. & Hizi, A. (2001) The ribonuclease H activity of the reverse transcriptases of human immunodeficiency viruses type 1 and type 2 is affected by the thumb subdomain of the small protein subunits. *J. Mol. Biol*. 311, 957–71.

Sevilya, Z., Loya, S., Adir, N. & Hizi, A. (2003) The ribonuclease H activity of the reverse transcriptases of human immunodeficiency viruses type 1 and type 2 is modulated by residue 294 of the small subunit. *Nucleic Acids Res.* 31, 1481–7.

Shah, F. S., Curr, K. A., Hamburgh, M. E., Parniak, M., Mitsuya, H., Arnez, J. G. & Prasad, V. R. (2000) Differential influence of nucleoside analog-resistance mutations K65R and L74V on the overall mutation rate and error specificity of human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem*. 275, 27037–44.

Shirasaka, T., Kavlick, M. F., Ueno, T., Gao, W.-Y., Kojima, E., Alcaide, M. L., Chokekijchai, S., Roy, B. M., Arnold, E., Yarchoan, R. & Mitsuya, H. (1995) Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. *Proc. Natl. Acad. Sci. U. S. A.* 92, 2398–402.

Skar, H., Borrego, P., Wallstrom, T. C., Mild, M., Marcelino, J. M., Barroso, H., Taveira, N., Leitner, T. & Albert, J. (2010) HIV-2 genetic evolution in patients with advanced disease is faster than that in matched HIV-1 patients. *J. Virol*. 84, 7412–5.

Smith, R. A., Anderson, D. J., Pyrak, C. L., Preston, B. D. & Gottlieb, G. S. (2009) Antiretroviral drug resistance in HIV-2: three amino acid changes are sufficient for classwide nucleoside analogue resistance. *J. Infect. Dis*. 199, 1323–6.

Smyth, R. P., Davenport, M. P. & Mak, J. (2012) The origin of genetic diversity in HIV-1. *Virus Res.* 169, 415–29.

Soll, L. & Berg, P. (1969) Recessive lethals: a new class of nonsense suppressors in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 63, 392–9.

Stopar, D., Spruijt, R. B., Wolfs, C. J. & Hemminga, M. A. (2003) Protein-lipid interactions of bacteriophage M13 major coat protein. *Biochim. Biophys. Acta*. 1611, 5–15.

Stuke, A. W., Ahmad-Omar, O., Hoefer, K., Hunsmann, G. & Jentsch, K. D. (1997) Mutations in the SIV env and the M13 lacZa gene generated in vitro by reverse transcriptases and DNA polymerases. *Arch. Virol.* 142, 1139–54.

Seelamgari, A., Maddukuri, A., Berro, R., de la Fuente, C., Kehn, K., Deng, L., Dadgar, S., Bottazzi, M. E., Ghedin, E., Pumfery, A. & Kashanchi, F. (2004) Role of viral regulatory and accessory proteins in HIV-1 replication. *Front. Biosci*. 9, 2388–413.

Strebel, K. (2013) HIV accessory proteins versus host restriction factors. *Curr. Opin. Virol*. 3, 692–9.

Sultana, S., Solotchi, M., Ramachandran, A. & Patel, S. S. (2017) Transcriptional fidelities of human mitochondrial POLRMT, yeast mitochondrial Rpo41, and phage T7 single-subunit RNA polymerases. *J. Biol. Chem.* 292, 18145−60.

Suzuki, Y. & Craigie, R. (2007) The road to chromatin - nuclear entry of retroviruses. *Nat. Rev. Microbiol.* 5, 187–96.

Svarovskaia, E. S., Cheslock, S. R., Zhang, W. H., Hu, W. S. & Pathak, V. K. (2003) Retroviral mutation rates and reverse transcriptase fidelity. *Front. Biosci*. 8, d117–34.

Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–26.

Tang, G. Q., Anand, V. S. & Patel, S. S. (2011) Fluorescence-based assay to measure the real-time kinetics of nucleotide incorporation during transcription elongation. *J. Mol. Biol.* 405, 666–78.

Taube, R., Avidan, O. & Hizi, A. (1997) The fidelity of misinsertion and mispair extension throughout DNA synthesis exhibited by mutants of the reverse transcriptase of human immunodeficiency virus type 2 resistant to nucleoside analogs. *Eur. J. Biochem*. 250, 106–114.

Telesnitsky, A. & Wolin, S. L. (2016) The host RNAs in retroviral particles. *Viruses* 8, E235.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22, 4673–80.

Tian, L., Kim, M. S., Li, H., Wang, J. & Yang, W. (2018) Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid. *Proc. Natl. Acad. Sci. U. S. A.* 115, 507−12.

Traverse, C. C. & Ochman, H. (2016) Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3311–6. Erratum in: 113, E4257–8 (2016).

Tunitskaya, V. L. & Kochetkov, S. N. (2002) Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry (Mosc.)* 67, 1124–35.

Ulrich, S. & Kool, E. T. (2011) Importance of steric effects on the efficiency and fidelity of transcription by T7 RNA polymerase. *Biochemistry* 50, 10343–49.

Visseaux, B., Damond, F., Matheron, S., Descamps, D. & Charpentier, C. (2016) HIV-2 molecular epidemiology. *Infect. Genet. Evol*. 46, 233–40. Erratum in: 58, 294 (2018).

Weber, J. & Grosse, E. (1989) Fidelity of human immunodeficiency virus type I reverse transcriptase in copying natural DNA. *Nucleic Acids Res*. 17, 1379–93.

Weiss, K. K., Chen, R., Skasko, M., Reynolds, H. M., Lee, K., Bambara, R. A., Mansky, L. M. & Kim, B. (2004) A role for dNTP binding of human immunodeficiency virus type 1 reverse transcriptase in viral mutagenesis. *Biochemistry* 43, 4490–500.

Weymouth, L. A. & Loeb, L. A. (1978) Mutagenesis during in vitro DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1924–8.

Yahi, N., Tamalet, C., Tourrès, C., Tivoli, N., Ariasi, F., Volot, F., Gastaut, J. A., Gallais, H., Moreau, J. & Fantini, J. (1999) Mutation patterns of the reverse transcriptase and protease genes in human immunodeficiency virus type 1-infected patients undergoing combination therapy: survey of 787 sequences. *J. Clin. Microbiol*. 37, 4099–106.

Yasukawa, K., Iida, K., Okano, H., Hidese, R., Baba, M., Yanagihara, I., Kojima, K., Takita, T. & Fujiwara, S. (2017) Next-generation sequencing-based analysis of reverse transcriptase fidelity. *Biochem. Biophys. Res. Commun*. 492, 147−53.

Yu, H. & Goodman, M. F. (1992) Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates. *J. Biol. Chem.* 267, 10888–96.

Yu, X. F., Ito, S., Essex, M. & Lee, T. H. (1988) A naturally immunogenic virion-associated protein specific for HIV-2 and SIV. *Nature* 335, 262–5.

Zakharova, O. D., Tarrago-Litvak, L., Fournier, M., Andreola, M. L., Repkova, M. N., Venyaminova, A. G., Litvak, S. & Nevinsky, G. A. (1995) Interaction of primer $tRNA^{Lys3}$ with the p51 subunit of human immunodeficiency virus type 1 reverse transcriptase: a possible role in enzyme activation. *FEBS Lett.* 361, 287–90.

Zhan, P., Liu, X. & De Clercq, E. (2010) Blocking nuclear import of pre-integration complex: an emerging anti-HIV-1 drug discovery paradigm. *Curr. Med. Chem*. 17, 495–503.

Zhou, S., Jones, C., Mieczkowski, P. & Swanstrom, R. (2015) Primer ID validates template sampling depth and greatly reduces the error rate of next-generation sequencing of HIV-1 genomic RNA populations. *J. Virol*. 89, 8540–55.